US 20040018533A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2004/0018533 A1**

Adam et al. (43) **Pub. Date: Jan. 29, 2004**

(54) **DIAGNOSING PREDISPOSITION TO FAT DEPOSITION AND THERAPEUTIC METHODS FOR REDUCING FAT DEPOSITION AND TREATMENT OF ASSOCIATED CONDITIONS**

(76) Inventors: **Gail Isabel Reid Adam**, Knivsta (SE); **Maria L. Langdown**, La Jolla, CA (US); **Richard B. Roth**, La Jolla, CA (US); **Mikhail F. Denissenko**, Poway, CA (US); **Kevin J. Smylie**, San Diego, CA (US)

Correspondence Address:
MORRISON & FOERSTER LLP
3811 VALLEY CENTRE DRIVE
SUITE 500
SAN DIEGO, CA 92130-2332 (US)

(21) Appl. No.: **10/455,552**

(22) Filed: **Jun. 4, 2003**

Related U.S. Application Data

(60) Provisional application No. 60/386,012, filed on Jun. 4, 2002.

Publication Classification

(51) **Int. Cl.**[7] **C12Q 1/68**

(52) **U.S. Cl.** **435/6**

(57) **ABSTRACT**

Provided herein are methods for prognosing and diagnosing fat deposition and related disorders (e.g., obesity and non-insulin diabetes dependent mellitus (NIDDM)) in a subject, reagents and kits for carrying out the methods, methods for identifying candidate therapeutics for reducing fat deposition and related disorders, and therapeutic methods for reducing fat deposition or treating fat deposition related disorders in a subject. These embodiments are based in part upon an analysis of polymorphic variations of the nucleic acid set forth in SEQ ID NO:1.

```
   1 aggatcagct tctttggtca ggcctaatga tgttaccaac tagtaattta taaaaacttg
  61 catttctaca ttttttcttaa ctgaccccat tccaaaacac aagcaaacca tctaaattca
 121 agatttgtga ggttgggact aactcattac tatagggtaa gtatgaatga aggctggaaa
 181 tgtgttttga ttgctgccat tatttttat ttttatttt tctgagatgg agtttcactc
 241 ttgtcgccca ggctggagtg caatggcgtg atcagctcat ggcaacctcc acctccccgg
 301 ttcaagcaat tctcctgcct cagcctcctg agtagctgag attacaggca tgcgccacca
 361 cgcccagcta atttttgtat ttttagtaga gacggggttt ctccattgtt ggtcaggctg
 421 gtctcgaact cccaatgtca ggtgatccac ccaccttggc ctgccaaagt gctgggatta
 481 aaggtgtaag ccacgtgcct ggtgtttttt ggggtttgtt tgtttgtttt gagatggagt
 541 ctgtcaccca ggctggagtg cagtggcaca atttcagctc actgcaacct ccgcctccca
 601 ggttcaagca attcttcctg cctcagcctc ccgagtagct gggattacag gcacctgcca
 661 taaggcccag ctaagttttt tatattttca gtagagatga ggtttcacca tgttggccag
 721 gctggtctcc agctcctgac ctcacatgat catccgcctg cctcggcctc ccaaagtgct
 781 gcagttatag gtgtgagcca ccgtgcccag ccatgttttg attcttaagg aaatgatttt
 841 taagtcccct tccaatgagg atttagcaat ggagttttct tcttctacag ataatttgag
 901 gcttcccaat tgcaccaaac tccatttggg cagatacctc acctgtccac tccatagccc
 961 tagcatcaag aaaagtgcca agcacacaac aggcactcag taaatttctg tcaaatgaat
1021 gaataaaata ccaccacctt ttcctgagac cattttggac caaggctggc atgggcatct
1081 ctgttcctcc ctctaactta agaattccat ctcttagagg cagatgctgc ctgttttctg
1141 caaattttat tattcacctt tctatatttg gaaaaacaag ttggtgtggc ttacttagga
1201 atctcatatc caggaaatga gctaaaaatg ttgtatcaaa gattatgcat aggatccaca
1261 aactaagaag caagcagatg agagcggtta gagacatcgt ggtatatttt ccggaagaaa
1321 tggtccatgg attaaaaatt gtcattatga tgagtctgca gtaacacaga aaccacgtag
1381 tgcgaatcag aaaatgaaaa gcacagccat tgactatgat tacaaaccaa aactttatac
1441 caattacctg ttggaattta tgctgcatga aggcaggata ttttgcttgt cctgttcact
1501 ggacagtact tggcataaag gatgtgctta tgaatatttt ctgggaaact aaatataaca
1561 ttcacctgtg aaaacacatc tgggccaggc gcagtggctc acgcctgtaa tcccaacact
1621 ttgggagccc gaggcgggtg gatcatggga ggtcaggagc tggagatcag cctggccaac
1681 atggtgaaaa cctgtctcta ataaaaatac aaaaattagc tgggcgtggt ggcgggcact
1741 tgtaatccca gctactcaag aggctgaggc aggagaattg cttgaacctg ggaggcggag
1801 gttacagtga gccgagatcg cgccactgca ctccagactg gacaagagtg aaactccgtc
1861 tcaaaaaaaa aaagaaaaag aaaaagaaaa aacgtatctg aagggaatga acaaatacac
1921 atgcggaaaa ttgtattcat ggggtagaac ggtggctgtt tccccccctt tcactatttt
1981 ctctcctaaa gttgtgatag cattttctca gtggaaagga tgaccacaga attatctccc
2041 ctgggctgca actaaggaag tatttggagt gagaatgcga ccaggcttca aatccaggag
2101 gggtatttta cagggcgtgt ggctctcacc cagggctcac aggacacctg atatttaggc
2161 tgttccccaa gaggctaagg acagcgcggc cccagctccc ccaaccacag ctgcccaagt
2221 ttaggagccc gcactcgtga gcgatggcta ctccgcacag ccccggggtc ctctacggaa
2281 atcccctggca cctatcagta cccctaaggg aggtcgctcg ttacccctca cttgaactct
2341 gtcccaaatg ggaggtcact ctctcttcgc ggaccccatc ttccccgaaa gggtctcttg
```

FIGURE 1A

2401 ttctctctgg ccctcctgag aaggtcatta ctccccaact cgccttttt gttgttttt
2461 ggtttctggt ttattgaggc agggtctcgc ttcatctccc agactggagt gcagtggcgc
2521 gatcacagct cacacggcag cctctacctc ctgggctcaa gcgatcctct cacctcagcc
2581 tccctgggac tacaggcgag cgccactacg cccgcttttt ttttttttt ttttttttt
2641 ttgtattttt gatagagacg aggtctcgct atgttgtcca ggctggtctc taactcaaga
2701 ttcctctcac ctcggcctcc gcaagtgctg ggatgacagg tgtgagccac cgcccccggc
2761 ccctcgcccg cctttgaag gagcctttcg tcctcaaggg cgaggccact ccccccccgc
2821 gagttccatg ccccctagag ggtcatcgtt cccgacgggg aggtggcgcc ctccccGGG
2881 CCCCGGGCCC Cgaccgcccg tgctgcctcc ttccgggccc tcctccgcga tgacggcgcc
2941 gccagcaggc caYgcggact gggcgggggt ccgagcgggg actgggaccc agaccgacta
3001 *ggggactggg agcgggcggc gcggccatgg cgggctgctg cKccgcgctg gcggccttcc*
3061 *tgttcgagta cgacacgccg cgcatcgtgc tcatccgcag ccgcaaagtg gggctcatga*
3121 *accgcgccgt gcaactgctc atcctggcct acgtcatcgg* gtgagcgtgg ggccgcgcgg
3181 ggggcgcggc gggtgctgcc ctcgcgtccg cgccgtgcgg cggctcatcc tggcctcggt
3241 cacctgggcg agtccgggag cggcgagccg aggcggtgac accttccctg ggccccagcc
3301 gccgcgccgg ggccccgggg gcgggaggct gcttgctgct gttttaaaac cacagcctgg
3361 gccaggcgcg gtggcgcaag cctgtaatcc gagcactttg ggagcccgag gcgggaccat
3421 cgcctgagcc caggagttcg aaaccaattt gggcaacatc gtgggacccc gtctctgcaa
3481 aaaattaaaa atcatccggg cctggtggcg cgcgcctggg gtcccagcta ctcaggaggc
3541 tgaggtggga ggatcgcttg agcccaagag gctgaggctg cagtgagcta tgatcacacc
3601 acttaactcg agcctgggag acagtccgag acccgtctgt ttacaaaata aataaaggca
3661 aagcctggct aggctgttac cctctcctgc ctgacccccc atagtgcccc ttggctgtag
3721 gggcttcgcc ctgaacgtct caatcttcgc tctccacttc ccatcctgct gatttgctga
3781 caaaaccaaa gcttggaatt tccgcttggt tccttctggg tgggtggctt ttggtcattt
3841 cttttgatca ctatgcggtg tcactatgtg gtagtagcga ggtcagactg tagcgagtgt
3901 ttaaagtttg cttcctttgt tttctgggct tgtggggctt tttgtggtac ctgccctagc
3961 ctagtcagtc attccccatg ctgccccctt aggctagaga tgccctaccg ccctcaggcc
4021 tcgctgaatg tgccattgta cttgaaggca cctgttactt ttttaatttt tttatttttt
4081 tgagacggaa tttctctctt gtcgcccggg ctggagtgca atggcacgat ctcggctcac
4141 tgcaacctct gcctcccggg ttcaagcaat tctcctgtct cagtctcctg agtggctggg
4201 attacaagtg cccgccacca cgcccggcca atttttgtat tttttagtag agacggggtt
4261 tcaggctggt cctgaactga cctcaggtga tccgcccacc tcagcctccc aaaatgctgg
4321 gattacaggc gtgagccccc acgcggggct ggcacctgtt acaattgaat acaacgcgag
4381 agagaagatg ataattacct tgccacctgg ggccacttta aaggcttacc tggggtgttc
4441 tggtgtctac agctgctggg ctgacacgta ttgcctagtg cagtggtctc caaactgcac
4501 attggaacca tctggaaatc tttggctccc atccccagac cctttttat ttttttgatg
4561 gagtctccct ctgtcgccca ggctggagtg cagtggtgca atcttggctc gctgcaacct
4621 ccacctccct ggttcaagcg attctcctgc ctcagcctcc tgagtagctg ggattacagg
4681 cacccaccac catgcccggg ctaatttttg tatttttagt agggacaggg tttcgccatg
4741 ttggccaggc tggtctcgaa ctcctggcct caggtgatcc acccgccgca gcctcccaga

FIGURE 1B

4801 gtgctgggat tacaggcgtg agccaccgcg cccagacgca tccccagacc ttctgattta
4861 attgctgcag agtatgatcg gctgcccagg tgattctcat gttcagcaac gtttcggaac
4921 caatgaactg tgtttcccaa gtttacctga tcatggaaga ctcatctgag gaattgttta
4981 aaagtaacag atccccccgt ctttggggtg aggtttggaa tcttaacact gcccaaggga
5041 agcccaagca ggtttggctt ccaatatctg tcagcagcag cagtagatac agacagtatt
5101 caagaggatg gctttccagg tgcagactca acatcagaac cacctggccc acacccagat
5161 ttggactcat tggctcacac ctgtagtccc agtgctttgg gaagccaagg tgggaggatt
5221 gcttgaggcc aggagctcaa gaccagcctg gacacaagac cagcctggac aacatagcga
5281 gaccccatct ctaagagaaa aataaagttt aaaatggtaa attgtactga tggagcagtg
5341 caaaaaaaag aatttttttt ttttttttt ctgaacacgg tggctcatgc ctataatctc
5401 agcactttgg gaggccgagg cgggtggatc acctgaggtc aggagttcga gaccagcctg
5461 gccaacatgg taaaaccccg tctctactaa aagtacaaaa aattagccag gcctagtggc
5521 taatgcctgt aatcccggct acatgggagg ctgaggcagg agaatcgttt gaatccagga
5581 ggcggaggtt gcagtgagcc cagattgcgc cactgcactc cagcctgggc aacaagagca
5641 aaactccatc tcaaatatat atatatatat atatatatat atatatattt ttttttttt
5701 tttggagact cactgctagc cagttaactg aactcacatc cctggttgga cctgaagcat
5761 ttgggggcag gagctcccct gcagcttttt gcttttggcg tttgctgcta gtgggaggat
5821 cagtgcacat ctaccttagg gtacctgtaa aattacaact gcggaaaata tccatagtgg
5881 tgttgcactg tatggtagcc actagtaact tgcacctgtt gaacacttaa ccgttcctac
5941 tctacactga gatgtgctat aaccttaaaa tacacactgg atttcaaaga catagtacag
6001 aaaaagaggc tgggcacagt ggctcatgcc tgaaatcaca gtactttggg aggctgaggt
6061 gggaagatcg cttgagctca agagtttgag acaaacctgg gcaacatagc gagaccttgc
6121 ctctacaaaa aaatacaaaa attagcctgg catggtggca cacgcctata ggcctaccta
6181 tttgggaggt tgaggcagaa ggactgcttg aacctgggag gttgaggcca cagagagctg
6241 tgatcacacc accctccagc ctgggcaaca gagcaagaca ctgtctcaaa aaaacaaaga
6301 gtaccactat cacatgaaaa catttttat tgatgacatg ttaaaatatt ttggatatat
6361 tgggttacac agaatatatc attacaatta atttttcctg tctcttttt ctttttttct
6421 ttttaaacgg cgccttgctt tgtcacccag actggagtgc aatgcactgc aacctctgcc
6481 tcccaggctc aagtgattct accatctcag cctctggaat tatgggacta caggtgcaca
6541 ccaccatgac tggctaatta ttgtgttttt gtttgtttgt ttgtacagat gagaaatctc
6601 atctctacaa tcatgaggtt tccatgttgc ccaggctggt ctcgaactcc tgagctctag
6661 caatctgtct gcctcggcct cccaaagtgc tgggattaca ggcatgagcc accgtgccca
6721 gccctttcta tttcttttta cttttgaca tggctactag aaacttttaa attacagggt
6781 cttgccctgt ggtagctcac gcctgtaatc ccaatacttt gggaggctga ggcaagtgga
6841 tcccttgagc ttaggaattt tgagaccagc ctgggcaacg tagtgaaacc ccatctctac
6901 caaaaatacg aaaaattagc ttggcatagt ggcgcacacc tgtggaccca gctacttggg
6961 aggctgaggt gggaggattg ctgtagcctg ggaggcggag gttgcagtga gccaagattg
7021 tgccactgca cttcagcctg ggcgacagag cgagagcctg tctcaaaaaa aaaaatatgt
7081 ggctcgcgtt atgtttttat gggacagcgc agctctgtag gattccttat gacgcattca
7141 ctatctcggt catgagaatc tttaaagcag taaatgcttt gcttttgtcc gtcttgctat

FIGURE 1C

7201 aatcaaaact gactcaggac ctgcattaag gctcttttt tttttaacct catctcttaa
7261 ctttttgatg caactcatct tagtgaaatt gtggtctcct cctacagcaa ccctccctca
7321 aaaaaagacc gaacacRcaa ttctcatcta tctaactctg ttaacaatca tatgaggagg
7381 aaataatcac atgataatga gtggcatgaa accctgtgtc tctaaattaa gcaactcact
7441 gtacttgatg agtacatcag tcttcctgcc ccaagctttt gtgtgtcaga ctggaattaa
7501 atggaagctt acactgcagc tgaattccta gaagccctgg acttgtcagg ggaaaagccc
7561 atgggggaa agcccatctc taggctgggc aatgaacttg gagactgctc gatcacacac
7621 actcataagt aacaaaacca ttaatttctt ttctttttt tggtggaggg acagagtctc
7681 actcttgccc aggctggagt gcagtggcat gatctcagcc cactgcaatc tccacctccc
7741 gggttcaagt gattctcctg cctcagcctc ccgagtatgt gggactacag gcacccacca
7801 ccacacccag ctaatttgt gtatttttag tagagtcggg gtttcacaat gttggccagg
7861 ctggtcttgg actcctgacc tcaacaaaac cagtgaattt ctgaagtctc agcatctagg
7921 acttagaaga aggagaagcc tagcatttcc tgattctata aataacaaca ctgatagtct
7981 ctagtagatg tttaacagca gttcgaaggg gaggggttct gatttgtagt actggccaat
8041 ttccacggtg tcagtacttg ctctgtgacc aacgatgact cacaaacatt gttgtaaaca
8101 tttgacctac attaattctt ttaccctcac aagaatcctg tgaagatcta ctatcattcc
8161 cagattatag atgggaaaaa ctgaggcaca atgaggttaa gcaccttggc caaagttatg
8221 cctcgtgaat ggtggagctg agccatgaac caagacctct gcctccagtc tgtcctaacc
8281 cctataagct aatgcctgta gatggttttt ttttttttt ttttttgag acagagcctt
8341 gccctgtcgc ccaggctgga gtgcggtggc atcacagctc actgcagcct ccgtctccca
8401 ggctcaagtg atcctcccac ctcagcctcc caaatagccg ggatttttat tggacagcac
8461 agctctgtag gattccttat gacgtattca ctgtctcagt cacgagaatc tttaaagcag
8521 taaatgcttt gcttttgtcc atcttgctat aatcaaaact gaccatgccc ggctaatttt
8581 ttatttttg tagagatccc tatgttgccc aggcttgtct cgagctcttg ggcacaaatg
8641 atcttccccc cttgacctcc caaagttctg ggattacagg cgtgagccac tgtacccagc
8701 ccctttggat ttttaattgt tcaaactggt agtcatagtc ccacttcaga ggagagaacc
8761 atatatgtga ccagaattct tgggggaaaa gtaagaaact actcttaaaa gttcttaaa
8821 gggtagaatt ctctttttcta gaaagagtag cctagggttg tccaatcttt tggctttcgt
8881 gggccacact ggaaggagaa attatcttgg gccacacatt aaatatgcta acactaacaa
8941 tagctgatga gctaaaaaac aaacttttt tgatcacaca aaaaaaaatc tcataacgtt
9001 ttaagaaagt ttacgaattt gtgttgggcc tcattcaaag ccatcctggg ccgcatgcgg
9061 cctgtgtgct atgggttaga caagcttaga gtagagcata gtgggtgaga gaagaggctt
9121 cagatgcttc tggcttgagt tcaaatcctg actctgccac tcactagctg tgtgacctttt
9181 ggcaagtgac ttaacctctc tgtgtttcca tttcctcctc tgtaaaatgg aaataataat
9241 aatgatacct gtctcataaa gtggtcagga ttaaatgagt taatacatat aaagagctta
9301 gatcagtgcc tggtgcctat aaacactcca gtggtaactg ttataatttg tattatctct
9361 tggtgattcc cataggtcac aggcattgaa agtgctcaaa gggccaggtg cagtggctca
9421 cgcctgtaat cccagcactt tgggaggtcg aggtgggtgg atcacctgag gtcaggagtt
9481 cgagaccagc ctggccaaca tggtgaaacc ccgtctctac taaaaataca aaattagcca
9541 ggcgtgttgg cgcatgcctg taatcccagc tacttgggaa gctgaggcag gagaattgcc

FIGURE 1D

9601 tgaacccggg aggcggaggt tgcagtgagc caagatcaca ccattgcact ccagcttagg
9661 caacaagagc gaaacatcgt ctcaaaaaaa aaaaaagaga aaatgctcaa atgtaataaa
9721 gacagtattc atcgagcgtt ttctctgtgc catcagtgtg ctgagtgctt tacctgcacc
9781 atctcattta tttcccgcaa accctcctgt ggtttaggtg ctaccatccc catttggcag
9841 aagtggaaat ggagtcccct agaaggtacg tagcttgtag aaagtcatgt aattagtcaa
9901 catcgtactt ccagccactg gccagcagaa gcttccccat caattcaaat ccattgaatc
9961 ctaatggcaa tttaagaggc ctgttttatt ttataggtgg gtgtttgtgt gggaaaaggg
10021 ctaccaggaa actgactccg tggtcagctc cgttacgacc aaggtcaagg gcgtggctgt
10081 gaccaacact tctaaacttg gattccggat ctgggatgtg gcggattatg tgataccagc
10141 tcaggtgtgt ctcccactgt gtcttctgtc taacactgac accttgctct ttactgactt
10201 tgcacacttg attaatgatt gacgtggcct gagtcctgac atcaccaatg cccagaatga
10261 ttgttttagt ccaagaaaac ggtgcagcca gatctctttt cagcttagga tctaacagag
10321 tgtatccagg cagctgtggt tcctcagctt gaaagcatgg caggctgaaa gccctgaaag
10381 atgagaaaac aaatgtctta aatcctgata gaaattcaag tggaaatatc acctttccac
10441 attttgggtt tgcattattg tcccaaatac acgaggtaaa tcttttcaga cagtgtaaaa
10501 actttaggcc gggcacagta gctcacgtgt gtaatcccag cactttggga ggctgggggcg
10561 ggcagatacc ttgagcccag gagtttgaga ccagcctggg caatgtggtg aaaccccatc
10621 tctataaaaa aaaaaaaaaa gaaaaaaaaa gaaaagaatt agagaccatt agagaccaaa
10681 acccaaacaa gaaggaaaca gagacagcag cacaggactc catcgggagc gaaaggaagg
10741 gaagacattg tttactgctg catactagtc ttcaggctgt taatctgcaa agctccaagt
10801 taaacattga ctctttagtc ctggtaacag ctctgcacca cacagtcaaa cccgacatag
10861 ccctccagat ggcctatgtg ctaagggacc gagggcgacc agtagagata agctgccaac
10921 tcttggctca gagctcttgc cagggaaaac aagaactcca gaaactcttt cttcaggtgt
10981 gttcttcccc ctgccttctc aagagaaggg tcttgtgggg tccccagagW ggagaccgag
11041 ctaatgagcc ataaaagtca ggaggaaggg ccaggtgcgg tggctcatgc ctgtaatccc
11101 agcactttgg gaggccgagg tcggtgtatc atgaggtcaa gagattgaga ccatcctggc
11161 taacatggtg aaaccctgtc tactaaaaat acaaaaatta gtggggtgtg gtggtgggtg
11221 cctgtagtcc cagctactag ggaggctgag gcaggagaat ctcttgaacc tgggaggcgg
11281 aggttgtagt gagccgagat agtgccactg cactccagcc tggtgacaga gcaagactcc
11341 gtctcaaaaa aaaagatcaa gtggaacagg ggatttgcag ttggctggag ggggcagtcc
11401 tgttagcatg tacctgtctc agcagagaga cggttgcctc ctgtgcaggc atagccttgt
11461 cccggtgagc tctgggaatg tcatgctcgg tgtgtttact gggcagatga ggaagggcac
11521 aagccccgct tttgtccctt ggtggaaagg agcaccccac aactcagcct tccgtggtgg
11581 atctgggcgg gggccgtgtg agttccccag ggctttcctc atgagctttc acgacgtctg
11641 tgttctttgt agatggaaca cagtcatcta tttgcagccc cttggctcat ctccagactg
11701 aaattaccct taatttgaag tagataagtg ataggaaaca ggccagatgt ggtggttcat
11761 gcctgtaatc tcaacacttt tggaggccga ggcagaagga tcacttgagc ccaggagttt
11821 gaggctgaag tgaactgtga tcacgccact gcactccagc tagggcaaca gagcaagacc
11881 ccaactctaa aagagagaga aagaaggaga gataggaaac gaaaagtaga ccatactgga
11941 ttgttccaag tataactcag gagagtcgta ggagcggcct tctctaacac caaatgcaga

FIGURE 1E

12001 gtgtgtttgt acattgaaca cctgggaacg tgacaggtgg cgggagagag gctgcagctc
12061 atgccacatc ctcacgcctc ctcctgtggc cttctccttg tctcccaaca tagtttgccc
12121 taggcctttg gaggacgcca gctatttcta gagcttccac cacttgtcag accgtggggt
12181 tggggatgtc tacagactgg ttggtttctg agttgccctg atccccaaga tggctcccac
12241 tactgactgt tatcctcagg agtgtgaaat gctaaaagct ttgggcatct ccttttaaaa
12301 tgcctgggcc cagccaggcg cagtggctta cgcctgtagt cccagcactt ttggaggcca
12361 aggcaggagg atcacttgag ctcaggattc aagaccagcc tgggcaatat agtgagaccc
12421 catcttgatt ggatgggtgg tggatggatg gatggatgga tggatggatg gatggatgga
12481 tagatagata gatagataga tagatagata gatagataga tggaaagaaa gaaagaatag
12541 ataaatgata ggataggaga aagaaaggga aagaaagaaa agagaagaga aaagaaagaa
12601 aagggatagg atagatgaca gatgatagga tagattagat aggtgataga taatcgataa
12661 atagaaaaca gatgatagat ataggataga ttagatagat agatagatag atagatagat
12721 agatagacag ataatagata gattagatta gatggatgga caccagcccc caatgagagg
12781 aggatccaac tagaatcttt tggggtcttt gccttttatt ccttatatga ggaggtcatt
12841 atcagcaatt cttcagtctc atccctagca aattgctttc cagaaactct aaatcaatta
12901 tttggcacct ttatacatat aaggaggaaa cagaaccttt tcaaagagaa atgtcacttg
12961 ctgaccctga ccgagtcttc tctagaatgt ttgatttata atttcagtta ggaaagaaaa
13021 ccagccagat ttgagcccca tgaggcctca tgtttttatt cctctgagat aatctggttg
13081 gcagggaatg tttttagcac ttgttggctc tttccctcag taagtcaagt ttgattgaga
13141 aaccattgtt tcagtttgca gacccccccc ttttttttaac tgcagtcaaa ttcacttata
13201 gaaactaacc attttaagtc gggcacagtg gctcatgctt gtaatcccag cactttggga
13261 ggctgaggca ggtagatcac ttgaggccag gagttggaga ccagcctagc caacatggtg
13321 aaatcctatc tctactaaag atacaaaaat tagccaggtg tggtgatgta cacctgtagt
13381 cccagctact caggagggtg aggctggaga atcgcttgaa cccaggaggc ggagattgca
13441 gtaagccaag atcgtgccac tgccctgcag cctgggcgac agagcaaggc tctgtctcaa
13501 ataaataaat aaataagaaa gaaattaacc attttaaaat gaaaaattca gcacatttag
13561 tgcattcaca gtgttatgca agcatcacct ctatctagtt caaaacattt ccatcacccc
13621 aaagggaagc ctcgtatgtt acacaaccac tccccattcc ccacttctct cactccctgg
13681 caacccgtaa gtgctttgtc tctatggact tgactgttct ggaaatttca catcaatgta
13741 atggtatcct gtgtggcctt tgtgtctggc ttctttcact cagcatcatg tttttggggt
13801 tcatccatgt tgtagcctgt atcagtgctt tattccttta ttttatttct tgagatgaaa
13861 tctcactctg tcacccaggc tggagtgcac tggcgtgatc tcggctcact gcaacctcca
13921 cctcccaggt tccagtgatt ctcctggctc agcccccctga gtagctggga ttacaggtgc
13981 ccgccaccac acccggctcc ttttgtgtgt gcgtgtgtct gtgtgtgttt gtgtgtgtgt
14041 gtgtgtgtgt gtgtgtgtgt ttttagtaca aagtttcacc atgttggcca ggctaatctg
14101 gaattcttga cctcaggtga tcctcctgcc tcagcctccc aaagtgctgg aattacaggt
14161 gtgagccacc atgcctggac ccttcattcc tttttattca tggctgaatc acattccatc
14221 atgtgactag actaggacag acatttgggc tgtttccacc attttttttt ttttttgag
14281 acagagtctc gtgtcatcca ggctggagtg cagtggcgcg atgtcggctc gctgcaacct
14341 ccgactcccg ggttcaagca attctcctgc ctcagcctcc caagtagctg ggactacagg

FIGURE 1F

14401 tgcccgccat cacacccggc taatttttgt atttttagta gagacggggg tttcactgtg
14461 ttgaccaggc tggtcttgaa ctcctgacct cgtgatccgc ccgcctcagc ctcccaaagt
14521 gctgggatta caggcttgag ccaccgcgcc cgacctgttt ccacctttg gctcttgtga
14581 aaagtgttgc tctgaatatt tatgtacaag gacttgtttg agtcactgtt ttccattggg
14641 ttataaactt tgatgggaga gagcgggaga gcacgcggtc agtgtttgag ttgtcctctt
14701 cagtcagcgt ctgccttttct tggctctccg tgagtcctct gagYgtggct tgcccgtgct
14761 gtctcccctc tgcaggagga aaactccctc ttcgtcatga ccaacgtgat cctcaccatg
14821 aaccagacac agggcctgtg ccccgaggta ggaggccccc gggaagagcc ccaggcccca
14881 cacccctctc cacgcctgtc cacctgtgtg tggggccggg ccacgtggac tttcttttcg
14941 ctccttcttt ttccagattc cagatgcgac cactgtgtgt aaatcagatg ccagctgtac
15001 tgccggctct gccggcaccc acagcaacgg tacgagcttg tggcctcctg gggagggcRg
15061 ccctgagca gatcgccccc actgtggagc gtctctgata gagaaatctt cccaattcct
15121 tcacatgacc ctgggtgagc caggtgccga ggctggggtc ctggagcccc tctacattca
15181 ctgctgtcat tggagcccca caagccatcc cagctcttgc cctactgttt tttttttgt
15241 tgttgttgtt gtttttttctt gtttttttttt ttgttttgtt tttttgtttt tttgagccag
15301 agtctcgctc tgtcacccaa actggagtgc aatggcatga tctcggctca ctgcaacctc
15361 cacctcccag tttcaaacaa ttctcctgcc tcagcctccc gagtagctgg gattacaggc
15421 atgcaccacc acgcccagct aatttttgta cttttggtag agatgggttt taccatgttg
15481 gccaggctgg tctcgaactc ctgacctcag gtgatccacc cgcctcagcc tcccaaagtg
15541 ctaggattac aggcatgacc cactgtgcct ggcctcttgt cccattcttt agcttggcat
15601 caccctggct gagatgtggc tggcacacag gtaactgtct tcctccgatt ctaactcctg
15661 gacagtgaca ctgtctaaat cccagatttg cacagctcaa gactggctgc atgggaggct
15721 ggatggggct ctcactccct actccaaaaa ggtagaaaat aggagacccc tgtggacatg
15781 ggacccccct gccaccttg tgcttgtagg agtctcaaca ggcaggtgcg tagctttcaa
15841 cgggtcYgtc aagacgtgtg aggtggcggc ctggtgcccg gtggaggatg acacacacgt
15901 gccacagtga gtccagccct agggaaggaa gtgccttttt gttttgtttt gttttagaca
15961 cagtttcact ctgtatcgcc aggttggagt gcagtggtgc katcttggct cacsacamcc
16021 tccacctccm gtgttcaagt gattttctgt gccwcagcct cccgagtagc tgagattata
16081 ggcacccacc accacgccca actaattttt tgtatttta gtagagacgg ggttttgcca
16141 tgttggccac gctggtctcc aactcctgac ctcaagtgat ccacctgcct tggcctccca
16201 aagtgctagg attacaggcg ggagccacca cacccagcca gaagccacca cacccagccg
16261 gaagcacctt caactctgcg tgtcagcctt gaacagaggt ggtcaaggag ttcacctgt
16321 gtgttgtggc tttttgtgtg tttctggtgt ctgtatgaag tcaggctgcc ctaggatggc
16381 tgctctcaac cacaaggcag cgaccatcag tggcaagatg ggccacaagt agtctgtcac
16441 tcacgagaag ggtgcaggtg actgacttcc ttgtatgtta ggactctgtc agttgcaaat
16501 aactgacagg aaacagttct cagtgccttа agcacatgtg cacaaacaca tacacacata
16561 cacatgcgtg cacagcgggg gggcagtgta ttgtccggct gaccaaaaag atccagcgtg
16621 gccaggagaa tctcgtgcca ggctgaggca gaagaatctc ttgaacctgg gaggaggagg
16681 ttgctgagct gagatggtgc cactgcactc cctcctgggc gacagaggcc actgcactcc
16741 tgcctgggcg acagagtgag agacaaagga caacacccta gtctcccttc ctggctctgc

FIGURE 1G

16801 atcccgccgt gccagcttca ctcgcaggac ccacagcata gcccgggagg cagcgctgca
16861 cttacagccc cccaagtctg agtctccctt tccagctgtc cctggagatt ctagacctgt
16921 ccatcatcgt atcaatggcc atatctctca acccatctct gtggctgggg aatgggtttt
16981 gcttatcagc ttgggcctgg gtcacatgcc cacctctgca gctggggcta gagtcagacc
17041 cacccaagtg acataggcaa gcattgggga gatatggctt ccaaaggaca tttgggttc
17101 tgttaccaga aaagctggga gcagatgctg ggaagaaaac aacggatgtt tgctacagtg
17161 tattttaaac taaactagac aggttcatcc tggcgcggtg gctcatgctg taatcctaac
17221 actttgggag gctgaggcag tcggatcatc tgaggccagg agttcaagac cagcctggcc
17281 aacatggaga aaccccgtct ccactaaaaa tacaaaaatt agggttaggg tggcgctygt
17341 ctgtaatccc agctacttgg gaggctgagg taggagaatt gcttgaaccc gggaagtgga
17401 ggttgcagtg agctgagatc gctcttctgc actctagctt gggcgacaga gagagactct
17461 gtctcaaaaa ataaaaattt aaaaaaataa ataaataaaa atttaaaaat caactaaact
17521 agccaggttc aaggttgctc tagaataaca ttcttcgtgg ttgtaattgg ctgtgcttca
17581 tggggtggag tccgggcagc gctctgagaa gtgcgtgcag cctagggcct ttctcaccga
17641 gagctcagac cgaggtctcc atgagccccg gggagcagga agccactgaa cctcatcctg
17701 ggtgtgttga gacagaagga aaggctgaga actgcagatt tagagattgc taaggcaaag
17761 gaaggagata gaaaaagaaa aggaatccta gcatcttaga gctggagagg atctcacaga
17821 gctgtgctct ccaacacagg agactctggc cacgcgtggc tattgaaaat tgaaatatgg
17881 ctgggtccga aaggagatgg gctgtccata ggaagtgcac ttatcatcaa attggatttt
17941 ttttttttt ttttgagacg aagtctcgct ctgtcgccca ggctggagta cagtggcgtg
18001 atcttggctc accacaacct ctgcctcccg ggttcaagtg attctcctgc ctcagcctcc
18061 tgagtagctg ggattatagg caccccccac cacaccttgc taatttttgt atttttagta
18121 gagatgcagg tttcaccatg ttggccaggc tggtctcgaa ctcctgacct caggttatcc
18181 acctgcctca gcctcccaaa gtgctgggat tataggtgtg agccactgca cccagcccga
18241 ttctgtttat tatttattta ttttttattt atttgtattt atttatttat ttgagatgga
18301 gtctcgctct gtcgcccagg ctggagtgca gtggcgccat ctctgttcac tgcaagctcc
18361 gcctcctgcg ctcacgccat tctcctgcct cagcctcccc agtagctggg actacaggcg
18421 cccaccacca cacccagcta atttttttgt atttttagta gagagggggt ttcactgtgt
18481 tagccaggat ggtctcgatc tcctgacctc gtgatccacc tgccttggcc tcccaaagtg
18541 ctgggattgc acgcgtgagt caccacgccc ggccctgaat atttattatg aaaaaaagaa
18601 tgtataatat ctcattgatt tttgtatgtt ggttacatgt tgaagtaata ttttgggtta
18661 ctgtataaaa ttttacttgt ttcttttact ttgtagtgtg gctattagga aactttaagt
18721 tggagtgtgg cttgtagttg tgtctcataa gatacttttt ttttgacaca ggatcttgct
18781 ctgtcaccca gtgcagtgat aacgatcatg gctcactgca gcttcgaact cctggactca
18841 agtgatcttc ccgcctcatc ctcctgtgta gctgggacca caggggcgca ccaccatgcc
18901 cagccaactt atttttttgta gagacagggt ctcaccatgt tgcccaggct ggtcttgaac
18961 tcctgggctc aagcaaccca cctgccttag cctcccaaag tgctgggatt acagcgtgag
19021 ccaccacacc tggcctcaca atatactttt attggacatc actgttgcag atctaatgcg
19081 gtgtaaaatg atttttaagt gcacactgat ataagaaata acattggcca ggcacagtgg
19141 ctcatgcctr taatccyarc actttgggag gcwgagrcrg kyggatcayt tgaggccagg

FIGURE 1H

19201 agttcaaawy cagcctggcc aacatggtga aacccccctct ctactaaaaa tataaaaatt
19261 agccaggcgt ggtggtgcgc acctgtaatc ccactactca ggaggctgag gcacgagaat
19321 tgcttgaacc cgggagatag aagttgcagt gagccaagat catgccactg cgctccagcc
19381 cgggtgacag aataagaccc tgtctcaaaa acagaacaaa gactaaaacg ttttggaact
19441 agacagaggt gatgattgca caacattatg aatgtactaa atgccacaga attgtccagt
19501 ttaaaatggt taatttttatg ttatgtgaat ttcatctcaa tttctctaaa aaggtgagtt
19561 gatcttaaag tattacatgt gaggtacaca ggtgtgtgtg tgggtgttac ggtacaggaa
19621 atgacaaaaa cttgaacacc cctgttgtaa gccatccctc ttgagggagt ggggactttg
19681 aagacctgag agaagtccag gcacagtggc ccatgcctgt aatcccagca ctttgggaga
19741 ccgaggtggg tggatcatct gaggtcagga gtttgagacc agcctgacca acatggtgaa
19801 accccatctc tactaaaaat atgaaaaaat tagccaggcg tggtggtgca cgcctgtaac
19861 cccagctact cgggaggctg aggcaggaga atcacttgaa cccagaggtt gcagtgagcc
19921 gagatygcgg cattgcactc cagcctgagc aacaggagca agactccatc tcaaaaaaaa
19981 aaaaaacctg agagaagccc cttggtccca gcctttctct gacagcagtg gtgacaggct
20041 cagctctcct ccgagtgcag ccctgtcact grccttgctc ctgtctcagg gctgggaaga
20101 ctgttgatgt tgtcattcca aagatcccac ctggatcagg ggaacatccc ccacagaagg
20161 gttagccata cagtgccaga ttctccagga gaaattcacc aaagaaatgg agtccccttg
20221 gggacagatt caacttgtat tgtcagccag gagctgacgt ggcacttctg agaagaggcc
20281 ggcgcacctg ctggcgggtg ctttgtgcac ttttcagaca ggtcaggatc cagcctgtag
20341 gcaaatttac ttttgctttg acctgtaaaa ccggatctgc ccagccttca ttctytccct
20401 ggagaacgcc tgcggcccca aagccaggcc tactgatttc cagtgaggcc acaaatcccc
20461 tccctggtta gcaattcagt tttcaccgcc ttgggggaga gggcctgccc ttccttgagg
20521 ggaagaacct ggaaggctcg gctcagtgtc ttctcttgga agagaagagt gtgcatgcag
20581 aagggtgtag aaaatgctag acgtgtttca tcttcttgac aaaatgacat tgtaagatgt
20641 gtgtatatgt ttttaaaaat attacatagg ggctgaacat ggtggcccac tcctgtagtc
20701 cagcacaggg aggcggaggt aggaagattg cttgaggtca ggagttcaag accagcctgg
20761 gcaatgtagg aagacttcat gtctacaaaa aaaaaaaaaa aaaatttcat tagccaggca
20821 tggtgatgca tgcctatagt tccagctact tgagaggctg aggtgggaga acccccctgag
20881 cctgggaggt cgaggctaca agaagctgtg ttcrtgcaac tgtactccag cctgggcaac
20941 agagcaagac cctgtctcaa aaatatatat aggctgggcg tgatggctta tgcctgtaat
21001 cccagcactt tgggaggctg aagcaggcgg atcacttgag gccaggagtt ggagaccagc
21061 ctgaccaaca tggyggaacc ctgtctctac taaaaataca aaaaaagtta accaggcata
21121 gtggctcaca cctgtgtaat cccagctaca cgggaggctg aggtggcagt gagctgagat
21181 cgcgccattg cactcaagcc tgcactcgag gtgacagagt gactccatct caaaaaaaaa
21241 aaaaaaaawa tatatatata tatatayaya cacacacaca casacasaca cacacaayac
21301 aatacatata tatatgtgtg trtatatata tatataacat ggttatgtga gtatttgtat
21361 gtattttatt aacagtaatg ttatcattgt ttttcagacc tgctttttta aaggctgcag
21421 aaaacttcac tcttttggtt aagaacaaca tctggtatcc caaatttaat ttcagcaagt
21481 aagtggtggc caggtgtgtg agttcaccag ggtcttggag aaacttctgg ctcttctctc
21541 ttctctgagg ttttcgtcgc tctgattttc tgcttcctct cgactttagg aggaatatcc

FIGURE 1I

21601 ttcccaacat caccactact tacctcaagt cgtgcattta tgatgctaaa acagatccct
21661 tctgccccat attccgtctt ggcaaaatag tggagaacgc aggacacRgt ttccaggaca
21721 tggccgtgga ggtgggtgcg ggccctggct ctcctgaccc agccctggag gcgtctcgtg
21781 ccaggtgctg aggaaagcct tgccgtgtct ctgctgctca tccccaggga ggcatcatgg
21841 gcatccaggt caactgggac tgcaacctgg acagagccgc ctccctctgc ttgcccaggt
21901 actccttccg ccgcctcgat acacgggacg ttgagcacaa cgtatctcct ggctacaatt
21961 tcaggtgggc gtgagcttgg gcccctcgct catgttgtag ggggtgctgg tggctgcgta
22021 cgtgccagtg ggccgcccac tgaagaccag cactcaggca gcaccccaag ggcaggctgc
22081 cggtcccccg tccaaggcgg cgggaaggcc atgctgggaa aatgccctta gtggtgtcct
22141 gctccggggc catcccggcc cccgagaccc ctccttgccc tttcctgccg caagaaacat
22201 gttgagatgg ttcttagagc agccaggaga agctgggggc ttaagctttc cagcacctgc
22261 ctcagccatg acctccattc actgcctcaa ggagcggatg atcttgtgat cctccagtcc
22321 aaaggcctct ggggcctggc ccaggaattg gtttctcaaa ggttgaacct gtgccagaat
22381 ctccagagtg caggggacac aggtctgagg cccctgaaag gccttgctgg gttccctgct
22441 gcaaatgctg gcatctcggt gtcctggggc agatgtaagc gagtcccaca aacagtggct
22501 tgcaactaca ggacttgatt ctctcacagt tctcacagtt ctagaggcca caacctaaaa
22561 tccaggcatc agcagggctg tgctcgctct gaaggctctg gagagagtcc ttacttgtct
22621 ctcctggctt ctggtagccc acggtgttcc ttggcttgtg gatgcatcac tccagtctct
22681 gcctctcctg tcatgaggcc ttctgccttg taYgtgtctc tgtgtctttt cttcttgtaa
22741 ggatatcagt cactgaattt agggcccacc ctactctagt atgacctcat cttaacaaat
22801 tacgtctgcc aatatcccat ttccaaataa ggtcacactc ccaagcttca ggtagatgtg
22861 aatttttggg gacacactgt tcaacccact acagctgatg atctcatgat cagaatctat
22921 agccaggtaa atttccacgt acccttttgtt gcatttttca cagtccttaa aaaggaagcg
22981 ctccaggccg ggcgcagtga ctcacacctg taatcctagc actttgggag gccaaggcgg
23041 gcagatcacc tgaggccagc atggcgaaac cccatctcta ctaaaaaaaa tacaaaaatt
23101 agctgggcat ggyggcgcat gcctgtaatc ccagcaactt gggaggctga ggcaggagaa
23161 tcgcttgagc cagggaggca gaggttgcag tgagccgaga tcgcactgct gcactccagc
23221 ctgggcgaca aagcaagtct ccgtctcagg gggaaaaaaa agaggccctg caaggaccat
23281 ccctgggcgc ggtcagcggt gctttgggct gcacttgcct ccccctagtg gtgacgtgtg
23341 gaatccgatc catacgggat cctaagacct ctgtaattgt gggatctgct gtcctttga
23401 accttctggc tcactgagta aagcagatat ttgaattttt cttttctagc agttagattt
23461 gtgttattcc atctctactg tcacacactt cttgtggaaa cctcaagctt aggtcagagc
23521 cctaaaacgg aaagamctca ggacgcctat cccccaacct cgggcagcac caggccctgc
23581 ctgtgtctcc caggcctggc ccatggcgtg gaagttgtgg gccagataga gcatggagct
23641 agggtgcggg ccgaggctcc tgatggatcc ctggagaaca tggggcagca gatggtatgt
23701 cttgacggtt cttcaccaac catccatccc ctgctcatgc gtgtgygctc tctcgcgtgc
23761 tcgctctctc tctctctcat atatgtatat gtatatatgt gtatatatat atatatatag
23821 ccattatata tatagccatt atatatatat atatttttt ttyttcttct tttttaaggg
23881 agtgacttgc tctgtcaccc aggttggagt gcagtggtgc gatttcagct cactgcagcc
23941 tttgcctccc gggttcaaga cattctcctg cctcagccgc ctgagtagct gggattacag

FIGURE 1J

24001 gcgcccgcca ccacacccag ctaatttttt gtatttttat tagagacggg gtttctccac
24061 attggccagg gtggtctcga gctcctcacc tcaagtgatc cacctgcctc ggcctcccaa
24121 agtgccggaa ttacaggtgt gagccaccgc acccagccat atatttttaa atggtcttac
24181 tgagatgcaa ttcacatgcc ataccgttca crcatttgaa gtgcagaatt catagttttc
24241 actatattca cagataagtg caggcttcac cacagtcgat tttagaacat ttccgtcatc
24301 tcaaaaagaa accctgcacc cttctgtcat ccccccggccc ctggcaacag tggtctactt
24361 tctgtctctg cagatgtgcc tcttctgagc atttcatata atcaatatgt ggccttttgc
24421 gtctggctcc ttccccttag cctgatgttt tcaagcttca tccatgtcgt agcctgtgtc
24481 agtacctcgt ctcttttttat ggccaaatca tattctgctg tgttgacaga cgacgttttg
24541 ttcatctctt tatctgtgtg acgaacgagt ggttcttttt gttgttgttg ttgttgttga
24601 agactgtgtg cacagaatca atgtttcttt gaaaaaataa ataaatttaa aaagccctag
24661 gtcaagagat tgagaccrtc atggccaact tggtgaaacc ccgtctctac taaaaaatcc
24721 aaaaattagc tgggcatggt ggtgtgcgcc tgtagtccca gctactcggg aggctgaggc
24781 agaagaattg cttgaactgg ggaggtggag gttgcagtga gccaagattg gtgcccctac
24841 actccagcct ggcaacagag ggagactcca tctcaaaaaa aaaaaaaaaa aaaaagtccc
24901 tgcactgatg ctgtgttggg gctggcttag tcccctcctg caatgctggc tggtcacatt
24961 cttactgtta gatgttggag gtcagggtcc cttgagggag gcaggcagga tgtaccaggt
25021 gggatgttga gagcaaactg ttcaggttca ggagaggaca ctggtgctgg aggaggaggt
25081 ctccctgtgc ccctgtacct cgtgggcccc gcatcctcct gcttgcacag cccgcctcag
25141 ccaggagagg ccccgagccg ctccagcgtc cattcagccg gcagagtgga ccattcacct
25201 gtgccagctc cactctaacg ttctctcaca gggcccaagg tcctgcccca gccatctccc
25261 cctgatccat cctccttccc ctcaggtttg ccaagtacta cagagacctg gctggcaacg
25321 agcagcgcac gctcatcaag gcctatggca tccgcttcga catcattgtg tttgggaagg
25381 tagctcgccg ccactggctc ccctccgtca ctccctgcag ggacaagggg cctctccctg
25441 cccctgcaga aacacttttt ttctttttcg gtgtcttggc aggcagggaa atttgacatc
25501 atcccctacta tgatcaacat cggctctggc ctggcactgc taggcatggt gagtggttta
25561 ggccctgcct tcaccctcac ggtgaggtga gaccctgggc tggggtcctg gtcctggccc
25621 taggccctag acctcagatg tgtttctaaa cttgaccctc ctaccttctt tcccttggcc
25681 cccagccctc ctcccacctt cccttctcca agaccacccc cctcaggtcc cagccttctc
25741 ccaagagatg ggagtgcctt tccattccgg taaagattcc aggcttctca ggaaggggca
25801 cgcaaagaat aagatgggtt gatgggttgc aagcatcctg gctcactctc accctatgct
25861 aaactcaggc gaccgtgctg tgtgacatca tagtcctcta ctgcatgaag aaaagactct
25921 actatcggga gaagaaatat aaatatgtgg aagattacga gcaggtaggc ccctcctggc
25981 ccccagcagg cacaggcctc tcatctcttg ggtgtgggag ccctgggcgt gggcctgtct
26041 ggggaggccc ttctgcagag gctggcacca gtgtggcgtg gtgtccccgt taacccgggc
26101 agtcctgcca ctctcagcag ctgctccatc cctaggcccc tgtactagat tgactcttta
26161 aacccagcct cgtttcaatg arcgacatct caggttggtg atgataatgc atgctctgag
26221 aatgcctgtg ggcacacact acttcagtgc accttgcgga acaggaaggg ttgggttcca
26281 ggcctgggac caacttgaga acccctggcg gtgaagtccc aggagcgcac ctcccctccg
26341 cctgccacaa ggggtcccag gggcaccttg atctgcttgt gtccttcttt gcagggtctt

FIGURE 1K

```
26401 gctagtgagc tggaccagtg aggcctaccc cacacctggg ctctccacag ccccatcaaa
26461 gaacagagag gaggaggagg gagaaatggc caccacatca ccccagagaa atttctggaa
26521 tctgattgag tctccactcc acaagcactc agggttcccc agcagctcct gtgtgttgtg
26581 tgcaggatct gtttgcccac tcggcccagg aggtcagcag tctgttcttg gctgggtcaa
26641 ctctgctttt cccgcaacct ggggttgtcg ggggagcgct ggcccgacgc agtggcactg
26701 ctgtggcttt cagggctgga gctggctttg ctcagaagcc tcctgtctcc agctctctcc
26761 aggacaggcc cagtcctctg aggcacggcg gctctgttca agcactttat gcggcagggg
26821 aggccgcctg gctgcagtca ctagacttgt agcaggcctg ggctgcaggc ttcccccga
26881 ccattccctg cagccatgcg gcagagctgg catttctcct cagagaagcg ctgtgctaag
26941 gtgatcgagg accagacatt aaagcgtgat tttcttaatc cctgtctgtt gtctcatagc
27001 atgtgctaga actttccttc ctacccttt acgacaacca gtaaatccac tattccggct
27061 cactgtgctc agaaaaggtc catgggatgg tctgtttctg gcacttatgc acattttccc
27121 ctacttcttt atttatttat tttatttttg aggcagagtc tcactctgtc gcccaggctg
27181 gagcgcagtg gtgcaatctc ggctcattgc aacctctgcc tcccgggttc aagcgattct
27241 cctgcctcag cctcccgagt agctgggatt acaggtgcat gccaccatgc ccggctaatt
27301 tttgtatttt tagtagagac agggtttcac cacattggcc agccaggctg gtctcgaact
27361 cctgacctca aatgatccay ctgccttagc ctctcaaagt gattggatta caggtgtgcg
27421 ccaccacacc tggcctattt attttttta ttacccaggc tggagtgcag tggcgcaatc
27481 actgctcact gcagtctcca cctcctgggc ttaagcgatc ctcctacctc agcctcccat
27541 agtgctggga ttacaagcat gagccactgc acccagctgc ctacttttta aagcctaata
27601 atttgtcatt agcttgtctt cctgcacagt tcacaggaca ctttttctct agtgacatct
27661 tacctgtgcc tccatttga cataaaatta caattctatg ctgtgggtaa cacaattttt
27721 ggtaacatgg aagtggcttc ctctcccttc cctgtttgct tcccaaaagt ctccaaccaa
27781 aagcagcagc acgcatacct agctccgcac aggttcgaaa tgttcatgaa gctatttcca
27841 gctgcagctt cacaccaact aaatgtttgt gttttttgt ttttgagaca gagtcttgcc
27901 ctattgccca ggctggagtg cagtggcatg atctcggctc actgcaacct ctgcctcctg
27961 ggttcaagca attctcctgc ctcagcctcc caagtagctg ggattacagg catctgccac
28021 cacacccagc taatttttgt atttttagta gagacagagg ttcaccatgt tggccaggct
28081 ggtctggaac tcctgacctc atgatccacc gcctcagcct cccaaagtgc tgggattaca
28141 ggcgtgagcc actgcaccca gccacacacc agctgaattc tgcatcaggt attcagaatc
28201 cccctagagg aagcagtttg cccttcgcct ctgccatgaa ttcacttcca agtactgaga
28261 ggaaccatta ttttacaaat actctgcttc tgctggataa atggccaacc tgggcttttc
28321 ttgcttttgg aaaaaaaaaa aaaaaaagtt taccaactcc ctttattgaa gctctggttc
28381 atgctgcctt ttaaaaaaaa atgcttgttg cctgtaatcc caggactttg ggaggccaag
28441 gtggatcayt taagcgtagg cattcaagac cagcccagtc aayatggtga gaccctatct
28501 ctacaaaaaa ttaaattagc cgggtgtggc ggtatatatc tgtagtccca gctacttggg
28561 tggctgaggc crgaggatca cttgagccca ggaggttgag gctgcagtga gctatg
```

FIGURE 1L

```
>NCBI b29     #1      aggatcagct tctttggtca ggcctaatga tgttaccaac tagtaatttta taaaaacttg
>NCBI old     #1      aggatcagct tctttggtca ggcctaatga tgttaccaac tagtaatttta taaaaacttg
                      ..........................................................
              #1      aggatcagct tctttggtca ggcctaatga tgttaccaac tagtaatttta taaaaacttg

>NCBI b29     #61     catttctaca ttttttttaa ctgaccccat tccaaaacac aagcaaacca tctaaattca
>NCBI old     #61     catttctaca tttttcttaa ctgaccccat tccaaaacac aagcaaacca tctaaattca
                      ..........................................................
              #61     catttctaca tttttcttaa ctgaccccat tccaaaacac aagcaaacca tctaaattca

>NCBI b29     #121    agatttgtga ggttgggact aactcattac tatagggtaa gtatgaatga aggctggaaa
>NCBI old     #121    agatttgtga ggttgggact aactcattac tatagggtaa gtatgaatga aggctggaaa
                      ..........................................................
              #121    agatttgtga ggttgggact aactcattac tatagggtaa gtatgaatga aggctggaaa

>NCBI b29     #181    tgtgtttttga ttgctgccat tatttttttat ttttattttt tctgagatgg agtttcactc
>NCBI old     #181    tgtgtttttga ttgctgccat tatttttttat ttttattttt tctgagatgg agtttcactc
                      ..........................................................
              #181    tgtgtttttga ttgctgccat tatttttttat ttttattttt tctgagatgg agtttcactc

>NCBI b29     #241    ttgtcgccca ggctggagtg caatggcgtg atcagctcat ggcaacctcc acctccccgg
>NCBI old     #241    ttgtcgccca ggctggagtg caatggcgtg atcagctcat ggcaacctcc acctccccgg
                      ..........................................................
              #241    ttgtcgccca ggctggagtg caatggcgtg atcagctcat ggcaacctcc acctccccgg

>NCBI b29     #301    ttcaagcaat tctcctgcct cagcctcctg agtagctgag attacaggca tgcgccacca
>NCBI old     #301    ttcaagcaat tctcctgcct cagcctcctg agtagctgag attacaggca tgcgccacca
                      ..........................................................
              #301    ttcaagcaat tctcctgcct cagcctcctg agtagctgag attacaggca tgcgccacca

>NCBI b29     #361    cgcccagcta atttttgtat ttttagtaga gacggggttt ctccattgtt ggtcaggctg
>NCBI old     #361    cgcccagcta atttttgtat ttttagtaga gacggggttt ctccattgtt ggtcaggctg
                      ..........................................................
              #361    cgcccagcta atttttgtat ttttagtaga gacggggttt ctccattgtt ggtcaggctg

>NCBI b29     #421    gtctcgaact cccaatgtca ggtgatccac ccaccttggc ctgccaaagt gctgggatta
>NCBI old     #421    gtctcgaact cccaatgtca ggtgatccac ccaccttggc ctgccaaagt gctgggatta
                      ..........................................................
              #421    gtctcgaact cccaatgtca ggtgatccac ccaccttggc ctgccaaagt gctgggatta

>NCBI b29     #481    aaggtgtaag ccacgtgcct ggtgtttttt ggggtttgtt tgtttgtttt gagatggagt
>NCBI old     #481    aaggtgtaag ccacgtgcct ggtgtttttt ggggtttgtt tgtttgtttt gagatggagt
                      ..........................................................
              #481    aaggtgtaag ccacgtgcct ggtgtttttt ggggtttgtt tgtttgtttt gagatggagt

>NCBI b29     #541    ctgtcaccca ggctggagtg cagtggcaca atttcagctc actgcaacct ccgcctccca
>NCBI old     #541    ctgtcaccca ggctggagtg cagtggcaca atttcagctc actgcaacct ccgcctccca
                      ..........................................................
              #541    ctgtcaccca ggctggagtg cagtggcaca atttcagctc actgcaacct ccgcctccca

>NCBI b29     #601    ggttcaagca attcttcctg cctcagcctc ccgagtagct gggattacag gcacctgcca
>NCBI old     #601    ggttcaagca attcttcctg cctcagcctc ccgagtagct gggattacag gcacctgcca
                      ..........................................................
              #601    ggttcaagca attcttcctg cctcagcctc ccgagtagct gggattacag gcacctgcca

>NCBI b29     #661    taaggcccag ctaagttttt tatattttca gtagagatga ggtttcacca tgttggccag
>NCBI old     #661    taaggcccag ctaagttttt tatattttca gtagagatga ggtttcacca tgttggccag
                      ..........................................................
              #661    taaggcccag ctaagttttt tatattttca gtagagatga ggtttcacca tgttggccag

>NCBI b29     #721    gctggtctcc agctcctgac ctcacatgat catccgcctg cctcggcctc ccaaagtgct
>NCBI old     #721    gctggtctcc agctcctgac ctcacatgat catccgcctg cctcggcctc ccaaagtgct
                      ..........................................................
              #721    gctggtctcc agctcctgac ctcacatgat catccgcctg cctcggcctc ccaaagtgct

>NCBI b29     #781    gcagttatag gtgtgagcca ccgtgcccag ccatgttttg attcttaagg aaatgatttt
>NCBI old     #781    gcagttatag gtgtgagcca ccgtgcccag ccatgttttg attcttaagg aaatgatttt
                      ..........................................................
              #781    gcagttatag gtgtgagcca ccgtgcccag ccatgttttg attcttaagg aaatgatttt

>NCBI b29     #841    taagtcccct tccaatgagg atttagcaat ggagttttct tcttctacag ataatttgag
>NCBI old     #841    taagtcccct tccaatgagg atttagcaat ggagttttct tcttctacag ataatttgag
                      ..........................................................
              #841    taagtcccct tccaatgagg atttagcaat ggagttttct tcttctacag ataatttgag

>NCBI b29     #901    gcttcccaat tgcaccaaac tccatttggg cagatacctc acctgtccac tccatagccc
>NCBI old     #901    gcttcccaat tgcaccaaac tccatttggg cagatacctc acctgtccac tccatagccc
                      ..........................................................
              #901    gcttcccaat tgcaccaaac tccatttggg cagatacctc acctgtccac tccatagccc

>NCBI b29     #961    tagcatcaag aaaagtgcca agcacacaac aggcactcag taaatttctg tcaaatgaat
>NCBI old     #961    tagcatcaag aaaagtgcca agcacacaac aggcactcag taaatttctg tcaaatgaat
                      ..........................................................
              #961    tagcatcaag aaaagtgcca agcacacaac aggcactcag taaatttctg tcaaatgaat
```

FIGURE 2A

```
>NCBI b29     #1021    gaataaaata ccaccacctt ttcctgagac catttttqqac caaqqctqqc atgggcatct
>NCBI old     #1021    gaataaaata ccaccacctt ttcctgagac catttttqqac caaqqctqqc atgggcatct
                       ..........................................................
              #1021    gaataaaata ccaccacctt ttcctgagac catttttqqac caaqqctqqc atgggcatct

>NCBI b29     #1081    ctgttcctcc ctctaactta agaattccat ctcttagagg cagatgctgc ctqttttctq
>NCBI old     #1081    ctgttcctcc ctctaactta agaattccat ctcttagagg cagatgctgc ctqttttctq
                       ..........................................................
              #1081    ctgttcctcc ctctaactta agaattccat ctcttagagg cagatgctgc ctqttttctq

>NCBI b29     #1141    caaattttat tattcacctt tctatatttg gaaaaacaag ttggtgtggc ttacttagqa
>NCBI old     #1141    caaattttat tattcacctt tctatatttg gaaaaacaag ttggtgtggc ttacttagqa
                       ..........................................................
              #1141    caaattttat tattcacctt tctatatttg gaaaaacaag ttggtgtggc ttacttagga

>NCBI b29     #1201    atctcatatc caggaaatga gctaaaaatg ttgtatcaaa gattatgcat aggatccaca
>NCBI old     #1201    atctcatatc caggaaatga gctaaaaatg ttgtatcaaa gattatgcat aggatccaca
                       ..........................................................
              #1201    atctcatatc caggaaatga gctaaaaatg ttgtatcaaa gattatgcat aggatccaca

>NCBI b29     #1261    aactaagaag caagcagatg agagcggtta gagacatcgt ggtatatttt ccggaagaaa
>NCBI old     #1261    aactaagaag caagcagatg agagcggtta gagacatcgt ggtatatttt ccggaagaaa
                       ..........................................................
              #1261    aactaagaag caagcagatg agagcggtta gagacatcgt ggtatatttt ccggaagaaa

>NCBI b29     #1321    tggtccatgg attaaaaatt gtcattatga tgagtctgca gtaacacaga aaccacgtag
>NCBI old     #1321    tggtccatgg attaaaaatt gtcattatga tgagtctgca gtaacacaga aaccacgtag
                       ..........................................................
              #1321    tggtccatgg attaaaaatt gtcattatga tgagtctgca gtaacacaga aaccacgtag

>NCBI b29     #1381    tgcgaatcag aaaatgaaaa gcacagccat tgactatgat tacaaaccaa aactttatac
>NCBI old     #1381    tgcgaatcag aaaatgaaaa gcacagccat tgactatgat tacaaaccaa aactttatac
                       ..........................................................
              #1381    tgcgaatcag aaaatgaaaa gcacagccat tgactatgat tacaaaccaa aactttatac

>NCBI b29     #1441    caattacctg ttggaattta tgctgcatga aggcaggata ttttgcttgt cctgttcact
>NCBI old     #1441    caattacctg ttggaattta tgctgcatga aggcaggata ttttgcttgt cctgttcact
                       ..........................................................
              #1441    caattacctg ttggaattta tgctgcatga aggcaggata ttttgcttgt cctgttcact

>NCBI b29     #1501    ggacagtact tggcataaag gatgtgctta tgaatatttt ctgggaaact aaatataaca
>NCBI old     #1501    ggacagtact tggcataaag gatgtgctta tgaatatttt ctgggaaact aaatataaca
                       ..........................................................
              #1501    ggacagtact tggcataaag gatgtgctta tgaatatttt ctgggaaact aaatataaca

>NCBI b29     #1561    ttcacctgtg aaaacacatc tgggccaggc gcagtggctc acgcctgtaa tcccaacact
>NCBI old     #1561    ttcacctgtg aaaacacatc tgggccaggc gcagtggctc acgcctgtaa tcccaacact
                       ..........................................................
              #1561    ttcacctgtg aaaacacatc tgggccaggc gcagtggctc acgcctgtaa tcccaacact

>NCBI b29     #1621    ttgggagccc gaggcgggtg gatcatggga ggtcaggagc tggagatcag cctggccaac
>NCBI old     #1621    ttgggagccc gaggcgggtg gatcatggga ggtcaggagc tggagatcag cctggccaac
                       ..........................................................
              #1621    ttgggagccc gaggcgggtg gatcatggga ggtcaggagc tggagatcag cctggccaac

>NCBI b29     #1681    atggtgaaaa cctgtctcta ataaaaatac aaaaattagc tgggcgtggt ggcgggcact
>NCBI old     #1681    atggtgaaaa cctgtctcta ataaaaatac aaaaattagc tgggcgtggt ggcgggcact
                       ..........................................................
              #1681    atggtgaaaa cctgtctcta ataaaaatac aaaaattagc tgggcgtggt ggcgggcact

>NCBI b29     #1741    tgtaatccca gctactcaag aggctgaggc aggagaattg cttgaacctg ggaggcggag
>NCBI old     #1741    tgtaatccca gctactcaag aggctgaggc aggagaattg cttgaacctg ggaggcggag
                       ..........................................................
              #1741    tgtaatccca gctactcaag aggctgaggc aggagaattg cttgaacctg ggaggcggag

>NCBI b29     #1801    gttacagtga gccgagatcg cgccactgca ctccagactg gacaagagtg aaactccgtc
>NCBI old     #1801    gttacagtga gccgagatcg cgccactgca ctccagactg gacaagagtg aaactccgtc
                       ..........................................................
              #1801    gttacagtga gccgagatcg cgccactgca ctccagactg gacaagagtg aaactccgtc

>NCBI b29     #1861    tcaaaaaaaa aaagaaaaag aaaaagaaaa aacgtatctg aagggaatga acaaatacac
>NCBI old     #1861    tcaaaaaaaa aaagaaaaag aaaaagaaaa aacgtatctg aagggaatga acaaatacac
                       ..........................................................
              #1861    tcaaaaaaaa aaagaaaaag aaaaagaaaa aacgtatctg aagggaatga acaaatacac

>NCBI b29     #1921    atgcggaaaa ttgtattcat ggggtagaac ggtggctgtt tccccccctt tcactatttt
>NCBI old     #1921    atgcggaaaa ttgtattcat ggggtagaac ggtggctgtt tccccccctt tcactatttt
                       ..........................................................
              #1921    atgcggaaaa ttgtattcat ggggtagaac ggtggctgtt tccccccctt tcactatttt

>NCBI b29     #1981    ctctcctaaa gttgtgatag cattttctca gtggaaagga tgaccacaga attatctccc
>NCBI old     #1981    ctctcctaaa gttgtgatag cattttctca gtggaaagga tgaccacaga attatctccc
                       ..........................................................
```

FIGURE 2B

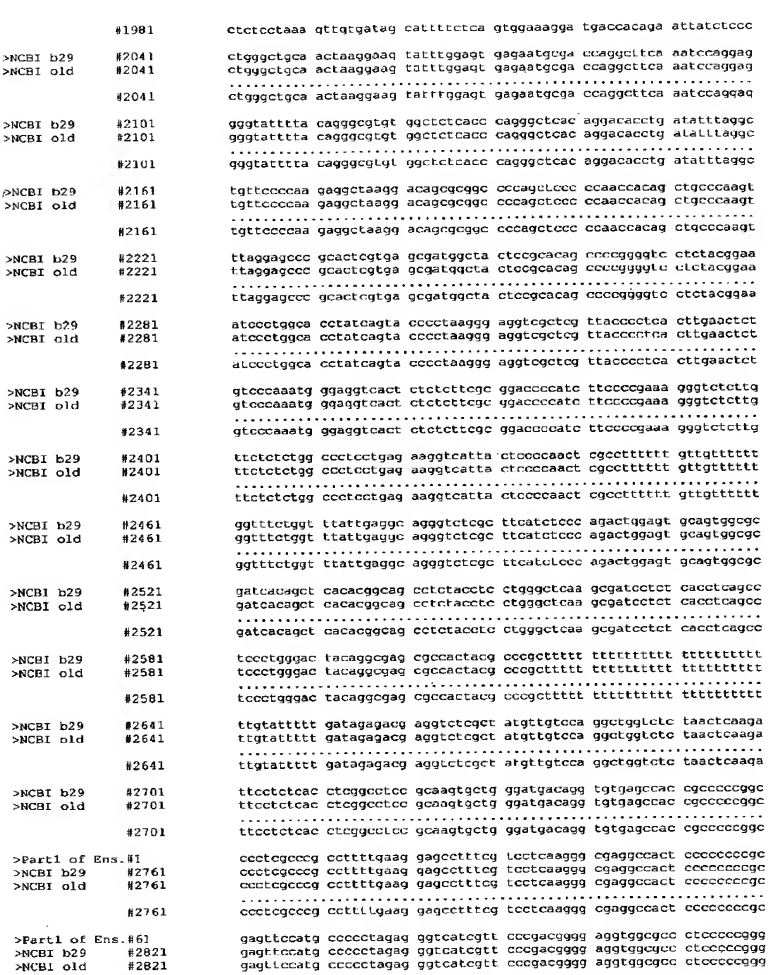


FIGURE 2C

```
                #2941       gccagcaggc caggcggact gggcggggct ccgagcgggg actgggaccc agaccgacta

>Part1 of Ens.#241          ggggactggg agcgggcggc gcggccatgg cgggctgctg cgccgcgctg gcggccttcc
>NCBI b29      #3001         ggggactggg agcgggcggc gcggccatgg cgggctgctg cgccgcgctg gcggccttcc
>NCBI old      #3001         ggggactggg agcgggcggc gcggccatgg cgggctgctg cgccgcgctg gcggccttcc
                            ..........................................................
                #3001       ggggactggg agcgggcggc gcggccatgg cgggctgctg cgccgcgctg gcggccttcc

>Part1 of Ens.#301          tgttcgagta cgacacgccg cgcatcgtgc tcatccgcag ccgcaaagtg gggctcatga
>NCBI b29      #3061         tgttcgagta cgacacgccg cgcatcgtgc tcatccgcag ccgcaaagtg gggctcatga
>NCBI old      #3061         tgttcgagta cgacacgccg cgcatcgtgc tcatccgcag ccgcaaagtg gggctcatga
                            ..........................................................
                #3061       tgttcgagta cgacacgccg cgcatcgtgc tcatccgcag ccgcaaagtg gggctcatga

>Part1 of Ens.#361          accgcgccgt gcaactgctc atcctggcct acgtcatcgg gtgagcgtgg ggccgccgcg
>NCBI b29      #3121         accgcgccgt gcaactgctc atcctggcct acgtcatcgg gtgagcgtgg ggccgccgcg
>NCBI old      #3121         accgcgccgt gcaactgctc atcctggcct acgtcatcgg gtgagcgtgg ggccgccgcg
                            ..........................................................
                #3121       accgcgccgt gcaactgctc atcctggcct acgtcatcgg gtgagcgtgg ggccgccgcg

>Part1 of Ens.#421          ggggccgcgg cgggtgctgcc ctcgcgtccg cgccgtgcgg cggctcatcc tggcctcggt
>NCBI b29      #3181         ggggccgcgg cgggtgctgcc ctcgcgtccg cgccgtgcgg cggctcatcc tggcctcggt
>NCBI old      #3181         ggggccgcgg cgggtgctgcc ctcgcgtccg cgccgtgcgg cggctcatcc tggcctcggt
                            ..........................................................
                #3181       ggggccgcgg cgggtgctgcc ctcgcgtccg cgccgtgcgg cggctcatcc tggcctcggt

>Part1 of Ens.#481          cacctgggcg agtccgggag cggcgagccg aggcggtgac accttccctg ggccccagcc
>NCBI b29      #3241         cacctgggcg agtccgggag cggcgagccg aggcggtgac accttccctg ggccccagcc
>NCBI old      #3241         cacctgggcg agtccgggag cggcgagccg aggcggtgac accttccctg ggccccagcc
                            ..........................................................
                #3241       cacctgggcg agtccgggag cggcgagccg aggcggtgac accttccctg ggccccagcc

>Part1 of Ens.#541          gccgcgccgg ggccccgggg gcgggaggct gcttgctgct gttttaaaac cacagcctgg
>NCBI b29      #3301         gccgcgccgg ggccccgggg gcgggaggct gcttgctgct gttttaaaac cacagcctgg
>NCBI old      #3301         gccgcgccgg ggccccgggg gcgggaggct gcttgctgct gttttaaaac cacagcctgg
                            ..........................................................
                #3301       gccgcgccgg ggccccgggg gcgggaggct gcttgctgct gttttaaaac cacagcctgg

>Part1 of Ens.#601          gccaggcgcg gtggcgcaag cctgtaatcc gagcactttg ggagcccgag gcgggaccat
>NCBI b29      #3361         gccaggcgcg gtggcgcaag cctgtaatcc gagcactttg ggagcccgag gcgggaccat
>NCBI old      #3361         gccaggcgcg gtggcgcaag cctgtaatcc gagcactttg ggagcccgag gcgggaccat
                            ..........................................................
                #3361       gccaggcgcg gtggcgcaag cctgtaatcc gagcactttg ggagcccgag gcgggaccat

>Part1 of Ens.#661          cgcctgagcc caggagttcg aaaccaattt gggcaacatc gtgggacccc gtctctgcaa
>NCBI b29      #3421         cgcctgagcc caggagttcg aaaccaattt gggcaacatc gtgggacccc gtctctgcaa
>NCBI old      #3421         cgcctgagcc caggagttcg aaaccaattt gggcaacatc gtgggacccc gtctctgcaa
                            ..........................................................
                #3421       cgcctgagcc caggagttcg aaaccaattt gggcaacatc gtgggacccc gtctctgcaa

>Part1 of Ens.#721          aaaattaaaa atcatccggg cctggtggcg cgcgcctggg gtcccagcta ctcaggaggc
>NCBI b29      #3481         aaaattaaaa atcatccggg cctggtggcg cgcgcctggg gtcccagcta ctcaggaggc
>NCBI old      #3481         aaaattaaaa atcatccggg cctggtggcg cgcgcctggg gtcccagcta ctcaggaggc
                            ..........................................................
                #3481       aaaattaaaa atcatccggg cctggtggcg cgcgcctggg gtcccagcta ctcaggaggc

>Part1 of Ens.#781          tgaggtggga ggatcgcttg agcccaagag gctgaggctg cagtgagcta tgatcacacc
>NCBI b29      #3541         tgaggtggga ggatcgcttg agcccaagag gctgaggctg cagtgagcta tgatcacacc
>NCBI old      #3541         tgaggtggga ggatcgcttg agcccaagag gctgaggctg cagtgagcta tgatcacacc
                            ..........................................................
                #3541       tgaggtggga ggatcgcttg agcccaagag gctgaggctg cagtgagcta tgatcacacc

>Part1 of Ens.#841          acttaactcg agcctgggag acagtccgag acccgtctgt ttacaaaata aataaaggca
>NCBI b29      #3601         acttaactcg agcctgggag acagtccgag acccgtctgt ttacaaaata aataaaggca
>NCBI old      #3601         acttaactcg agcctgggag acagtccgag acccgtctgt ttacaaaata aataaaggca
                            ..........................................................
                #3601       acttaactcg agcctgggag acagtccgag acccgtctgt ttacaaaata aataaaggca

>Part1 of Ens.#901          aagcctggct aggctgttac cctctcctgc ctgacccccc atagtgcccc ttggctgtag
>NCBI b29      #3661         aagcctggct aggctgttac cctctcctgc ctgacccccc atagtgcccc ttggctgtag
>NCBI old      #3661         aagcctggct aggctgttac cctctcctgc ctgacccccc atagtgcccc ttggctgtag
                            ..........................................................
                #3661       aagcctggct aggctgttac cctctcctgc ctgacccccc atagtgcccc ttggctgtag

>Part1 of Ens.#961          gggcttcgcc ctgaacgtct caatcttcgc tctccacttc ccatcctgct gatttgctga
>NCBI b29      #3721         gggcttcgcc ctgaacgtct caatcttcgc tctccacttc ccatcctgct gatttgctga
>NCBI old      #3721         gggcttcgcc ctgaacgtct caatcttcgc tctccacttc ccatcctgct gatttgctga
                            ..........................................................
                #3721       gggcttcgcc ctgaacgtct caatcttcgc tctccacttc ccatcctgct gatttgctga

>Part1 of Ens.#1021         caaaaccaaa gcttggaatt tccgcttggt tccttctggg tgggtggctt ttggtcattt
>NCBI b29      #3781         caaaaccaaa gcttggaatt tccgcttggt tccttctggg tgggtggctt ttggtcattt
>NCBI old      #3781         caaaaccaaa gcttggaatt tccgcttggt tccttctggg tgggtggctt ttggtcattt
                            ..........................................................
```

FIGURE 2D

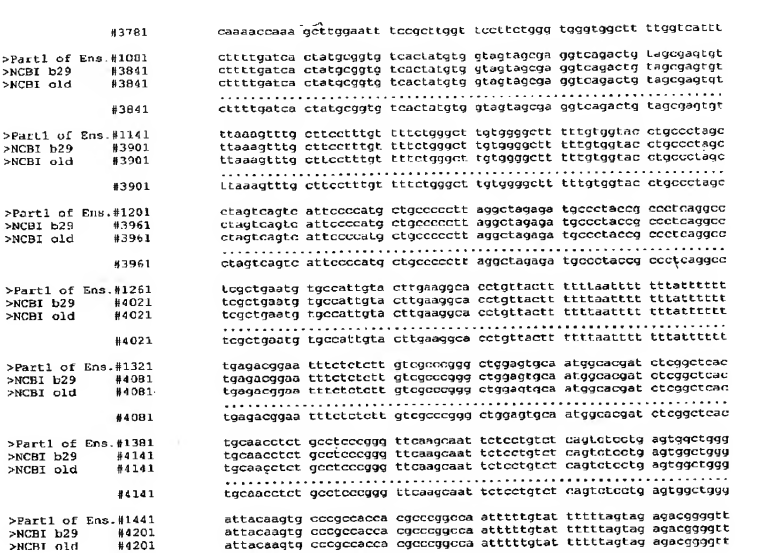


FIGURE 2E

```
             #4621   ccacctccct ggttcaagcg attctcctgc ctcagcctcc tgagtagctg ggattacagg

>Part1 of Ens.#1921  caccracrac catgcccggg ctaatttttg tatttttagt agggacaggg tttcgccatg
>NCBI b29      #4681  caccccaccac catgccc-gg ctaatttttg tatttttagt agggacaggg tttcgccatg
>NCBI old      #4681  caccccaccac catgcccggg ctaatttttg tatttttagt agggacaggg tttcgccatg
                      ............................................................
             #4681   caccccaccac catgcccggg ctaatttttg tatttttagt agggacaggg tttcgccatg
                                 *
>Part1 of Ens.#1981  ttggccaggc tggtctcgaa ctcctggcct caggtgatcc acccgccgca gcctcccaga
>NCBI b29      #4741  ttggccaggc tggtctcgaa ctcctggcct caggtgatcc acccgccgca gcctcccaga
>NCBI old      #4741  ttggccaggc tggtctcgaa ctcctggcct caggtgatcc acccgccgca gcctcccaga
                      ............................................................
             #4741   ttggccaggc tggtctcgaa ctcctggcct caggtgatcc acccgccgca gcctcccaga

>Part1 of Ens.#2041  gtgctgggat tacaggcgtg agccaccgcg cccagacgca tccccagacc ttctgattta
>NCBI b29      #4801  gtgctgggat tacaggcgtg agccaccgcg cccagacgca tccccagacc ttctgattta
>NCBI old      #4801  gtgctgggat tacaggcgtg agccaccgcg cccagacgca tccccagacc ttctgattta
                      ............................................................
             #4801   gtgctgggat tacaggcgtg agccaccgcg cccagacgca tccccagacc ttctgattta

>Part1 of Ens.#2101  attgctgcag agtatgatcg gctgcccagg tgattctcat gttcagcaac gtttcggaac
>NCBI b29      #4861  attgctgcag agtatgatcg gctgcccagg tgattctcat gttcagcaac gtttcggaac
>NCBI old      #4861  attgctgcag agtatgatcg gctgcccagg tgattctcat gttcagcaac gtttcggaac
                      ............................................................
             #4861   attgctgcag agtatgatcg gctgcccagg tgattctcat gttcagcaac gtttcggaac

>Part1 of Ens.#2161  caatgaactg tgtttcccaa gtttacctga tcatggaaga ctcatctgag gaattgttta
>NCBI b29      #4921  caatgaactg tgtttcccaa gtttacctga tcatggaaga ctcatctgag gaattgttta
>NCBI old      #4921  caatgaactg tgtttcccaa gtttacctga tcatggaaga ctcatctgag gaattgttta
                      ............................................................
             #4921   caatgaactg tgtttcccaa gtttacctga tcatggaaga ctcatctgag gaattgttta

>Part1 of Ens.#2221  aaagtaacag atccccccgt ctttggggtg aggtttggaa tcttaacact gcccaaggga
>NCBI b29      #4981  aaagtaacag atccccccgt ctttggggtg aggtttggaa tcttaacact gcccaaggga
>NCBI old      #4981  aaagtaacag atccccccgt ctttggggtg aggtttggaa tcttaacact gcccaaggga
                      ............................................................
             #4981   aaagtaacag atccccccgt ctttggggtg aggtttggaa tcttaacact gcccaaggga

>Part1 of Ens.#2281  agcccaagca ggtttggctt ccaatatctg tcagcagcag cagtagatac agacagtatt
>NCBI b29      #5041  agcccaagca ggtttggctt ccaatatctg tcagcagcag cagtagatac agacagtatt
>NCBI old      #5041  agcccaagca ggtttggctt ccaatatctg tcagcagcag cagtagatac agacagtatt
                      ............................................................
             #5041   agcccaagca ggtttggctt ccaatatctg tcagcagcag cagtagatac agacagtatt

>Part1 of Ens.#2341  caagaggatg gctttccagg tgcagactca acatcagaac cacctggccc acacccagat
>NCBI b29      #5101  caagaggatg gctttccagg tgcagactca acatcagaac cacctggccc acacccagat
>NCBI old      #5101  caagaggatg gctttccagg tgcagactca acatcagaac cacctggccc acacccagat
                      ............................................................
             #5101   caagaggatg gctttccagg tgcagactca acatcagaac cacctggccc acacccagat

>Part1 of Ens.#2401  ttggactcat tggctcacac ctgtagtccc agtgctttgg gaagccaagg tgggaggatt
>NCBI b29      #5161  ttggactcat tggctcacac ctgtagtccc agtgctttgg gaagccaagg tgggaggatt
>NCBI old      #5161  ttggactcat tggctcacac ctgtagtccc agtgctttgg gaagccaagg tgggaggatt
                      ............................................................
             #5161   ttggactcat tggctcacac ctgtagtccc agtgctttgg gaagccaagg tgggaggatt

>Part1 of Ens.#2461  gcttgaggcc aggagctcaa gaccagcctg gacacaagac cagcctggac aacatagcga
>NCBI b29      #5221  gcttgaggcc aggagctcaa gaccagcctg gacacaagac cagcctggac aacatagcga
>NCBI old      #5221  gcttgaggcc aggagctcaa gaccagcctg gacacaagac cagcctggac aacatagcga
                      ............................................................
             #5221   gcttgaggcc aggagctcaa gaccagcctg gacacaagac cagcctggac aacatagcga

>Part1 of Ens.#2521  gaccccatct ctaagagaaa aataaagttt aaaatggtaa attgtactga tggagcagtg
>NCBI b29      #5281  gaccccatct ctaagagaaa aataaagttt aaaatggtaa attgtactga tggagcagtg
>NCBI old      #5281  gaccccatct ctaagagaaa aataaagttt aaaatggtaa attgtactga tggagcagtg
                      ............................................................
             #5281   gaccccatct ctaagagaaa aataaagttt aaaatggtaa attgtactga tggagcagtg

>Part1 of Ens.#2581  caaaaaaaag aatttttttt tttttttttt ctgaacacgg tggctcatgc ctataatctc
>NCBI b29      #5341  caaaaaaaag aatttttttt tttttttttt ctgaacacgg tggctcatgc ctataatctc
>NCBI old      #5341  caaaaaaaag aatttttttt tttttttttt ctgaacacgg tggctcatgc ctataatctc
                      ............................................................
             #5341   caaaaaaaag aatttttttt tttttttttt ctgaacacgg tggctcatgc ctataatctc

>Part1 of Ens.#2641  agcactttgg gaggccgagg cgggtggatc acctgaggtc aggagttcga gaccagcctg
>NCBI b29      #5401  agcactttgg gaggccgagg cgggtggatc acctgaggtc aggagttcga gaccagcctg
>NCBI old      #5401  agcactttgg gaggccgagg cgggtggatc acctgaggtc aggagttcga gaccagcctg
                      ............................................................
             #5401   agcactttgg gaggccgagg cgggtggatc acctgaggtc aggagttcga gaccagcctg

>Part1 of Ens.#2701  gccaacatgg taaaaccccg tctctactaa aagtacaaaa aattagccag gcctagtggc
>NCBI b29      #5461  gccaacatgg taaaaccccg tctctactaa aagtacaaaa aattagccag gcctagtggc
>NCBI old      #5461  gccaacatgg taaaaccccg tctctactaa aagtacaaaa aattagccag gcctagtggc
                      ............................................................
```

FIGURE 2F

```
                #5461        gccaacatgg taaaacccca tctctactaa aagtacaaaa aattagccag gcctagtggc

>Part1 of Ens.#2761          taatgcctgt aatcccggct acatgggagg ctgaggcagg agaatcgttt gaatccagga
>NCBI b29      #5521         taatgcctgt aatcccggct acatgggagg ctgaggcagg agaatcgttt gaatccagga
>NCBI old      #5521         taatgcctgt aatcccggct acatgggagg ctgaggcagg agaatcgttt gaatccagga
                             ..........................................................
                #5521        taatgcctgt aatcccggct acatgggagg ctgaggcagg agaatcgttt gaatccagga

>Part1 of Ens.#2821          ggcggaggtt gcagtgagcc cagattgcgc cactgcactc cagcctgggc aacaagagca
>NCBI b29      #5581         ggcggaggtt gcagtgagcc cagattgcgc cactgcactc cagcctgggc aacaagagca
>NCBI old      #5581         ggcggaggtt gcagtgagcc cagattgcgc cactgcactc cagcctgggc aacaagagca
                             ..........................................................
                #5581        ggcggaggtt gcagtgagcc cagattgcgc cactgcactc cagcctgggc aacaagagca

>Part1 of Ens.#2881          aaactccatc tcaaatatat atatatatat atatatatat atatatattt ttttttttt
>NCBI b29      #5641         aaactccatc tcaaatatat atatatatat atatatatat atatatattt ttttttttt
>NCBI old      #5641         aaactccatc tcaaatatat atatatatat atatatatat atatatattt ttttttttt
                             ..........................................................
                #5641        aaactccatc tcaaatatat atatatatat atatatatat atatatattt ttttttttt

>Part1 of Ens.#2941          tttggagact cactgctagc cagttaactg aactcacatc cctggttgga cctgaagcat
>NCBI b29      #5701         tttggagact cactgctagc cagttaactg aactcacatc cctggttgga cctgaagcat
>NCBI old      #5701         tttggagact cactgctagc cagttaactg aactcacatc cctggttgga cctgaagcat
                             ..........................................................
                #5701        tttggagact cactgctagc cagttaactg aactcacatc cctggttgga cctgaagcat

>Part1 of Ens.#3001          ttgggggcag gagctcccct gcagcttttt gcttttggcg tttgctgcta gtgggaggat
>NCBI b29      #5761         ttgggggcag gagctcccct gcagcttttt gcttttggcg tttgctgcta gtgggaggat
>NCBI old      #5761         ttgggggcag gagctcccct gcagcttttt gcttttggcg tttgctgcta gtgggaggat
                             ..........................................................
                #5761        ttgggggcag gagctcccct gcagcttttt gcttttggcg tttgctgcta gtgggaggat

>Part1 of Ens.#3061          cagtgcacat ctaccttagg gtacctgtaa aattacaact gcggaaaata tccatagtgg
>NCBI b29      #5821         cagtgcacat ctaccttagg gtacctgtaa aattacaact gcggaaaata tccatagtgg
>NCBI old      #5821         cagtgcacat ctaccttagg gtacctgtaa aattacaact gcggaaaata tccatagtgg
                             ..........................................................
                #5821        cagtgcacat ctaccttagg gtacctgtaa aattacaact gcggaaaata tccatagtgg

>Part1 of Ens.#3121          tgttgcactg tatggtagcc actagtaact tgcacctgtt gaacacttaa ccgttcctac
>NCBI b29      #5881         tgttgcactg tatggtagcc actagtaact tgcacctgtt gaacacttaa ccgttcctac
>NCBI old      #5881         tgttgcactg tatggtagcc actagtaact tgcacctgtt gaacacttaa ccgttcctac
                             ..........................................................
                #5881        tgttgcactg tatggtagcc actagtaact tgcacctgtt gaacacttaa ccgttcctac

>Part1 of Ens.#3181          tctacactga gatgtgctat aaccttaaaa tacacactgg atttcaaaga catagtacag
>NCBI b29      #5941         tctacactga gatgtgctat aaccttaaaa tacacactgg atttcaaaga catagtacag
>NCBI old      #5941         tctacactga gatgtgctat aaccttaaaa tacacactgg atttcaaaga catagtacag
                             ..........................................................
                #5941        tctacactga gatgtgctat aaccttaaaa tacacactgg atttcaaaga catagtacag

>Part1 of Ens.#3241          aaaaagaggc tgggcacagt ggctcatgcc tgaaatcaca gtactttggg aggctgaggt
>NCBI b29      #6001         aaaaagaggc tgggcacagt ggctcatgcc tgaaatcaca gtactttggg aggctgaggt
>NCBI old      #6001         aaaaagaggc tgggcacagt ggctcatgcc tgaaatcaca gtactttggg aggctgaggt
                             ..........................................................
                #6001        aaaaagaggc tgggcacagt ggctcatgcc tgaaatcaca gtactttggg aggctgaggt

>Part1 of Ens.#3301          gggaagatcg cttgagctca agagtttgag acaaacctgg gcaacatagc gagaccttgc
>NCBI b29      #6061         gggaagatcg cttgagctca agagtttgag acaaacctgg gcaacatagc gagaccttgc
>NCBI old      #6061         gggaagatcg cttgagctca agagtttgag acaaacctgg gcaacatagc gagaccttgc
                             ..........................................................
                #6061        gggaagatcg cttgagctca agagtttgag acaaacctgg gcaacatagc gagaccttgc

>Part1 of Ens.#3361          ctctacaaaa aaatacaaaa attagcctgg catggtggca cacgcctata ggcctaccta
>NCBI b29      #6121         ctctacaaaa aaatacaaaa attagcctgg catggtggca cacgcctata ggcctaccta
>NCBI old      #6121         ctctacaaaa aaatacaaaa attagcctgg catggtggca cacgcctata ggcctaccta
                             ..........................................................
                #6121        ctctacaaaa aaatacaaaa attagcctgg catggtggca cacgcctata ggcctaccta

>Part1 of Ens.#3421          tttgggaggt tgaggcagaa ggactgcttg aacctgggag gttgaggcca cagagagctg
>NCBI b29      #6181         tttgggaggt tgaggcagaa ggactgcttg aacctgggag gttgaggcca cagagagctg
>NCBI old      #6181         tttgggaggt tgaggcagaa ggactgcttg aacctgggag gttgaggcca cagagagctg
                             ..........................................................
                #6181        tttgggaggt tgaggcagaa ggactgcttg aacctgggag gttgaggcca cagagagctg

>Part1 of Ens.#3481          tgatcacacc accctccagc ctgggcaaca gagcaagaca ctgtctcaaa aaaacaaaga
>NCBI b29      #6241         tgatcacacc accctccagc ctgggcaaca gagcaagaca ctgtctcaaa aaaacaaaga
>NCBI old      #6241         tgatcacacc accctccagc ctgggcaaca gagcaagaca ctgtctcaaa aaaacaaaga
                             ..........................................................
                #6241        tgatcacacc accctccagc ctgggcaaca gagcaagaca ctgtctcaaa aaaacaaaga

>Part1 of Ens.#3541          gtaccactat cacatgaaaa catttttttat tgatgacatg ttaaaatatt ttggatatat
>NCBI b29      #6301         gtaccactat cacatgaaaa catttttttat tgatgacatg ttaaaatatt ttggatatat
>NCBI old      #6301         gtaccactat cacatgaaaa catttttttat tgatgacatg ttaaaatatt ttggatatat
                             ..........................................................
```

FIGURE 2G

```
                #6301       gtaccactat cacatgaaaa catttttat tgatgacatg ttaaaatatt ttggaLatat

>Partl of Ens.#3601         tgggttacac agaatatatc attacaatta atttttcctg tctcttttt ctttttttct
>NCBI b29      #6361        tgggttacac agaatatatc attacaatta atttttcctg tctcttttt ctttttttct
>NCBI old      #6761        tgggttacac agaatatatc attacaatta atttttcctg tctcttttt ctttttttct
                            ..........................................................
                #6361       tgggttacac agaatatatc attacaatta atttttcctg tctcttttt ctttttttct

>Partl of Ens.#3661         ttttaaacgg cgccttgctt tgtcacccag actggagtgc aatgcactgc aacctctgcc
>NCBI b29      #6421        ttttaaacgg cgccttgctt tgtcacccag actggagtgc aatgcactgc aacctctgcc
>NCBI old      #6421        ttttaaacgg cgccttgctt tgtcacccag actggagtgc aatgcactgc aacctctgcc
                            ..........................................................
                #6421       ttttaaacgg cgccttgctt tgtcacccag actggagtgc aatgcactgc aacctctgcc

>Partl of Ens.#3721         tcccaggctc aagtgattct accatctcag cctctggaat tatgggacta caggtgcaca
>NCBI b29      #6481        tcccaggctc aagtgattct accatctcag cctctggaat tatgggacta caggtgcaca
>NCBI old      #6481        tcccaggctc aagtgattct accatctcag cctctggaat tatgggacta caggtgcaca
                            ..........................................................
                #6481       tcccaggctc aagtgattct accatctcag cctctggaat tatgggacta caggtgcaca

>Partl of Ens.#3781         ccaccatgac tggctaatta ttgtgttttt gtttgtttgt ttgtacagat gagaaatctc
>NCBI b29      #6541        ccaccatgac tggctaatta ttgtgttttt gtttgtttgt ttgtacagat gagaaatctc
>NCBI old      #6541        ccaccatgac tggctaatta ttgtgttttt gtttgtttgt ttgtacagat gagaaatctc
                            ..........................................................
                #6541       ccaccatgac tggctaatta ttgtgttttt gtttgtttgt ttgtacagat gagaaatctc

>Partl of Ens.#3841         atctctacaa tcatgaggtt tccatgttgc ccaggctggt ctcgaactcc tgagctctag
>NCBI b29      #6601        atctctacaa tcatgaggtt tccatgttgc ccaggctggt ctcgaactcc tgagctctag
>NCBI old      #6601        atctctacaa tcatgaggtt tccatgttgc ccaggctggt ctcgaactcc tgagctctag
                            ..........................................................
                #6601       atctctacaa tcatgaggtt tccatgttgc ccaggctggt ctcgaactcc tgagctctag

>Partl of Ens.#3901         caatctgtct gcctcggcct cccaaagtgc tgggattaca ggcatgagcc accgtgccca
>NCBI b29      #6661        caatctgtct gcctcggcct cccaaagtgc tgggattaca ggcatgagcc accgtgccca
>NCBI old      #6661        caatctgtct gcctcggcct cccaaagtgc tgggattaca ggcatgagcc accgtgccca
                            ..........................................................
                #6661       caatctgtct gcctcggcct cccaaagtgc tgggattaca ggcatgagcc accgtgccca

>Partl of Ens.#3961         qcccttttta tttcttttta ctttttgaca tggctactag aaacttttaa attacagggt
>NCBI b29      #6721        gcccttttta tttcttttta ctttttgaca tggctactag aaacttttaa attacagggt
>NCBI old      #6721        gcccttttta tttcttttta ctttttgaca tggctactag aaacttttaa attacagggt
                            ..........................................................
                #6721       gcccttttta tttcttttta ctttttgaca tggctactag aaacttttaa attacagggt

>Partl of Ens.#4021         cttgccctgt ggtagctcac gcctgtaatc ccaatacttt gggaggctga ggcaagtgga
>NCBI b29      #6781        cttgccctgt ggtagctcac gcctgtaatc ccaatacttt gggaggctga ggcaagtgga
>NCBI old      #6781        cttgccctgt ggtagctcac gcctgtaatc ccaatacttt gggaggctga ggcaagtgga
                            ..........................................................
                #6781       cttgccctgt ggtagctcac gcctgtaatc ccaatacttt gggaggctga ggcaagtgga

>Partl of Ens.#4081         tcccttgagc ttaggaattt tgagaccagc ctgggcaacg tagtgaaacc ccatctctac
>NCBI b29      #6841        tcccttgagc ttaggaattt tgagaccagc ctgggcaacg tagtgaaacc ccatctctac
>NCBI old      #6841        tcccttgagc ttaggaattt tgagaccagc ctgggcaacg tagtgaaacc ccatctctac
                            ..........................................................
                #6841       tcccttgagc ttaggaattt tgagaccagc ctgggcaacg tagtgaaacc ccatctctac

>Partl of Ens.#4141         caaaaatacg aaaaattagc ttggcatagt ggcgcacacc tgtggaccca gctacttggg
>NCBI b29      #6901        caaaaatacg aaaaattagc ttggcatagt ggcgcacacc tgtggaccca gctacttggg
>NCBI old      #6901        caaaaatacg aaaaattagc ttggcatagt ggcgcacacc tgtggaccca gctacttggg
                            ..........................................................
                #6901       caaaaatacg aaaaattagc ttggcatagt ggcgcacacc tgtggaccca gctacttggg

>Partl of Ens.#4201         aggctgaggt gggaggattg ctgtagcctg ggaggcggag gttgcagtga gccaagattg
>NCBI b29      #6961        aggctgaggt gggaggattg ctgtagcctg ggaggcggag gttgcagtga gccaagattg
>NCBI old      #6961        aggctgaggt gggaggattg ctgtagcctg ggaggcggag gttgcagtga gccaagattg
                            ..........................................................
                #6961       aggctgaggt gggaggattg ctgtagcctg ggaggcggag gttgcagtga gccaagattg

>Partl of Ens.#4261         tgccactgca cttcagcctg ggcgacagag cgagagcctg tctcaaaaaa aaaaatatgt
>NCBI b29      #7021        tgccactgca cttcagcctg ggcgacagag cgagagcctg tctcaaaaaa aaaaatatgt
>NCBI old      #7021        tgccactgca cttcagcctg ggcgacgag cgagagcctg tctcaaaaaa aaaaatatgt
                            ..........................................................
                #7021       tgccactgca cttcagcctg ggcgacagag cgagagcctg tctcaaaaaa aaaaatatgt

>Partl of Ens.#4321         ggctcgcgtt atgtttttat gggacagcgc agctctgtag gattccttat gacgcattca
>NCBI b29      #7081        ggctcgcgtt atgtttttat gggacagcgc agctctgtag gattccttat gacgcattca
>NCBI old      #7081        ggctcgcgtt atgtttttat gggacagcgc agctctgtag gattccttat gacgcattca
                            ..........................................................
                #7081       ggctcgcgtt atgtttttat gggacagcgc agctctgtag gattccttat gacgcattca

>Partl of Ens.#4381         ctatctcggt catgagaatc tttaaagcag taaatgcttt gcttttgtcc gtcttgctat
>NCBI b29      #7141        ctatctcggt catgagaatc tttaaagcag taaatgcttt gcttttgtcc gtcttgctat
>NCBI old      #7141        ctatctcggt catgagaatc tttaaagcag taaatgcttt gcttttgtcc gtcttgctat
                            ..........................................................
```

FIGURE 2H

```
                #7141       ctatctcggt catgagaatc tttaaagcag taaatgcttt gcttttgtcc gtcttgctat

>Part1 of Ens.#4441         aatcaaaact gactcaggac ctgcattaag gctcttttct ttttaacct catctcttaa
>NCBI b29      #7201        aatcaaaact gactcaggac ctgcattaag gctcttttct ttttaacct catctcttaa
>NCBI old      #7201        aatcaaaact gactcaggac ctgcattaag gctcttttct ttttaacct catctcttaa
                            ..........  ..........  ..........  ..........  ..........
                #7201       aatcaaaact gactcaggac ctgcattaag gctcttttct ttttaacct catctcttaa

>Part1 of Ens.#4501         cttttgatg caactcatct tagtgaaatt gtggtctcct cctacagcaa cctcccctca
>NCBI b29      #7261        cttttgatg caactcatct tagtgaaatt gtggtctcct cctacagcaa cctcccctca
>NCBI old      #7261        cttttgatg caactcatct tagtgaaatt gtggtctcct cctacagcaa cctcccctca
                            ..........  ..........  ..........  ..........  ..........
                #7261       cttttgatg caactcatct tagtgaaatt gtggtctcct cctacagcaa cctcccctca

>Part1 of Ens.#4561         aaaaaagacc gaacacgcaa ttctcatcta tctaactctg ttaacaatca tatgaggagg
>NCBI b29      #7321        aaaaaagacc gaacacgcaa ttctcatcta tctaactctg ttaacaatca tatgaggagg
>NCBI old      #7321        aaaaaagacc gaacacgcaa ttctcatcta tctaactctg ttaacaatca tatgaggagg
                            ..........  ..........  ..........  ..........  ..........
                #7321       aaaaaagacc gaacacgcaa ttctcatcta tctaactctg ttaacaatca tatgaggagg

>Part1 of Ens.#4621         aaataatcac atgataatga gtggcatgaa accctgtgtc tctaaattaa gcaactcact
>NCBI b29      #7381        aaataatcac atgataatga gtggcatgaa accctgtgtc tctaaattaa gcaactcact
>NCBI old      #7381        aaataatcac atgataatga gtggcatgaa accctgtgtc tctaaattaa gcaactcact
                            ..........  ..........  ..........  ..........  ..........
                #7381       aaataatcac atgataatga gtggcatgaa accctgtgtc tctaaattaa gcaactcact

>Part1 of Ens.#4681         gtacttgatg agtacatcag tcttcctgcc ccaagctttt gtgtgtcaga ctggaattaa
>NCBI b29      #7441        gtacttgatg agtacatcag tcttcctgcc ccaagctttt gtgtgtcaga ctggaattaa
>NCBI old      #7441        gtacttgatg agtacatcag tcttcctgcc ccaagctttt gtgtgtcaga ctggaattaa
                            ..........  ..........  ..........  ..........  ..........
                #7441       gtacttgatg agtacatcag tcttcctgcc ccaagctttt gtgtgtcaga ctggaattaa

>Part1 of Ens.#4741         atggaagctt acactgcagc tgaattccta gaagccctgg acttgtcagg ggaaaagccc
>NCBI b29      #7501        atggaagctt acactgcagc tgaattccta gaagccctgg acttgtcagg ggaaaagccc
>NCBI old      #7501        atggaagctt acactgcagc tgaattccta gaagccctgg acttgtcagg ggaaaagccc
                            ..........  ..........  ..........  ..........  ..........
                #7501       atggaagctt acactgcagc tgaattccta gaagccctgg acttgtcagg ggaaaagccc

>Part1 of Ens.#4801         atgggggggaa agcccatctc taggctgggc aatgaacttg gagactgctc gatcacacac
>NCBI b29      #7561        atgggggggaa agcccatctc taggctgggc aatgaacttg gagactgctc gatcacacac
>NCBI old      #7561        atgggggggaa agcccatctc taggctgggc aatgaacttg gagactgctc gatcacacac
                            ..........  ..........  ..........  ..........  ..........
                #7561       atgggggggaa agcccatctc taggctgggc aatgaacttg gagactgctc gatcacacac

>Part1 of Ens.#4861         actcataagt aacaaaacca ttaattttct ttctttttt tggtggaggg acagagtctc
>NCBI b29      #7621        actcataagt aacaaaacca ttaattttct ttctttttt tggtggaggg acagagtctc
>NCBI old      #7621        actcataagt aacaaaacca ttaattttct ttctttttt tggtggaggg acagagtctc
                            ..........  ..........  ..........  ..........  ..........
                #7621       actcataagt aacaaaacca ttaattttct ttctttttt tggtggaggg acagagtctc

>Part1 of Ens.#4921         actcttgccc aggctggagt gcagtggcat gatctcagcc cactgcaatc tccacctccc
>NCBI b29      #7681        actcttgccc aggctggagt gcagtggcat gatctcagcc cactgcaatc tccacctccc
>NCBI old      #7681        actcttgccc aggctggagt gcagtggcat gatctcagcc cactgcaatc tccacctccc
                            ..........  ..........  ..........  ..........  ..........
                #7681       actcttgccc aggctggagt gcagtggcat gatctcagcc cactgcaatc tccacctccc

>Part1 of Ens.#4981         gggttcaagt gattctcctg cctcagcctc ccgagtatgt gggactacag gcacccacca
>NCBI b29      #7741        gggttcaagt gattctcctg cctcagcctc ccgagtatgt gggactacag gcacccacca
>NCBI old      #7741        gggttcaagt gattctcctg cctcagcctc ccgagtatgt gggactacag gcacccacca
                            ..........  ..........  ..........  ..........  ..........
                #7741       gggttcaagt gattctcctg cctcagcctc ccgagtatgt gggactacag gcacccacca

>Part1 of Ens.#5041         ccacacccag ctaatttttgt gtatttttag tagagtcggg gtttcacaat gttggccagg
>NCBI b29      #7801        ccacacccag ctaatttttgt gtatttttag tagagtcggg gtttcacaat gttggccagg
>NCBI old      #7801        ccacacccag ctaatttttgt gtatttttag tagagtcggg gtttcacaat gttggccagg
                            ..........  ..........  ..........  ..........  ..........
                #7801       ccacacccag ctaatttttgt gtatttttag tagagtcggg gtttcacaat gttggccagg

>Part1 of Ens.#5101         ctggtcttgg actcctgacc tcaacaaaac tagtgaattt ctgaagtctc agcatctagg
>NCBI b29      #7861        ctggtcttgg actcctgacc tcaacaaaac tagtgaattt ctgaagtctc agcatctagg
>NCBI old      #7861        ctggtcttgg actcctgacc tcaacaaaac tagtgaattt ctgaagtctc agcatctagg
                            ..........  ..........  ..........  ..........  ..........
                #7861       ctggtcttgg actcctgacc tcaacaaaac tagtgaattt ctgaagtctc agcatctagg

>Part1 of Ens.#5161         acttagaaga aggagaagcc tagcatttcc tgattctata aataacaaca ctgatagtct
>NCBI b29      #7921        acttagaaga aggagaagcc tagcatttcc tgattctata aataacaaca ctgatagtct
>NCBI old      #7921        acttagaaga aggagaagcc tagcatttcc tgattctata aataacaaca ctgatagtct
                            ..........  ..........  ..........  ..........  ..........
                #7921       acttagaaga aggagaagcc tagcatttcc tgattctata aataacaaca ctgatagtct

>Part1 of Ens.#5221         ctagtagatg tttaacagca gttcgaaggg gaggggttct gatttgtagt actggccaat
>NCBI b29      #7981        ctagtagatg tttaacagca gttcgaaggg gaggggttct gatttgtagt actggccaat
>NCBI old      #7981        ctagtagatg tttaacagca gttcgaaggg gaggggttct gatttgtagt actggccaat
                            ..........  ..........  ..........  ..........  ..........
```

FIGURE 21

```
                #7981          ctagtagatg tttaacagca gttcgaaggg gaggggttct gatttgtagt actggccaat

>Part1 of Ens.#5281            ttccacggtg tcagtacttg ctctgtgacc aacgatgact cacaaacatt gttgtaaaca
>NCBI b29      #8041           ttccacggtg tcagtacttg ctctgtgacc aacgatgact cacaaacatt gttgtaaaca
>NCBI old      #8041           ttccacggtg tcagtacttg ctctgtgacc aacgatgact cacaaacatt gttgtaaaca
                               ..........................................................

                #8041          ttccacggtg tcagtacttg ctctgtgacc aacgatgact cacaaacatt gttgtaaaca

>Part1 of Ens.#5341            tttgacctac attaattctt ttaccctcac aagaatcctg tgaagatcta ctatcattcc
>NCBI b29      #8101           tttgacctac attaattctt ttaccctcac aagaatcctg tgaagatcta ctatcattcc
>NCBI old      #8101           tttgacctac attaattctt ttaccctcac aagaatcctg tgaagatcta ctatcattcc
                               ..........................................................

                #8101          tttgacctac attaattctt ttaccctcac aagaatcctg tgaagatcta ctatcattcc

>Part1 of Ens.#5401            cagattatag atgggaaaaa ctgaggcaca atgaggttaa gcaccttggc caaagttatg
>NCBI b29      #8161           cagattatag atgggaaaaa ctgaggcaca atgaggttaa gcaccttggc caaagttatg
>NCBI old      #8161           cagattatag atgggaaaaa ctgaggcaca atgaggttaa gcaccttggc caaagttatg
                               ..........................................................

                #8161          cagattatag atgggaaaaa ctgaggcaca atgaggttaa gcaccttggc caaagttatg

>Part1 of Ens.#5461            cctcgtgaat ggtggagctg agccatgaac caagacctct gcctccagtc tgtcctaacc
>NCBI b29      #8221           cctcgtgaat ggtggagctg agccatgaac caagacctct gcctccagtc tgtcctaacc
>NCBI old      #8221           cctcgtgaat ggtggagctg agccatgaac caagacctct gcctccagtc tgtcctaacc
                               ..........................................................

                #8221          cctcgtgaat ggtggagctg agccatgaac caagacctct gcctccagtc tgtcctaacc

>Part1 of Ens.#5521            cctataagct aatgcctgta gatggttttt ttttttttt ttttttgag acagagcctt
>NCBI b29      #8281           cctataagct aatgcctgta gatggttttt tttttttttt ttttttgag acagagcctt
>NCBI old      #8281           cctataagct aatgcctgta gatggttttt tttttttttt ttttttgag acagagcctt
                               ..........................................................

                #8281          cctataagct aatgcctgta gatggttttt tttttttttt ttttttgag acagagcctt

>Part1 of Ens.#5581            gccctgtcgc ccaggctgga gtgcggtggc atcacagctc actgcagcct ccgtctccca
>NCBI b29      #8341           gccctgtcgc ccaggctgga gtgcggtggc atcacagctc actgcagcct ccgtctccca
>NCBI old      #8341           gccctgtcgc ccaggctgga gtgcggtggc atcacagctc actgcagcct ccgtctccca
                               ..........................................................

                #8341          gccctgtcgc ccaggctgga gtgcggtggc atcacagctc actgcagcct ccgtctccca

>Part1 of Ens.#5641            ggctcaagtg atcctcccac ctcagcctcc caaatagccg ggatttttat tggacagcac
>NCBI b29      #8401           ggctcaagtg atcctcccac ctcagcctcc caaatagccg ggatttttat tggacagcac
>NCBI old      #8401           ggctcaagtg atcctcccac ctcagcctcc caaatagccg ggatttttat tggacagcac
                               ..........................................................

                #8401          ggctcaagtg atcctcccac ctcagcctcc caaatagccg ggatttttat tggacagcac

>Part1 of Ens.#5701            agctctgtag gattccttat gacgtattca ctgtctcagt cacgagaatc tttaaagcag
>NCBI b29      #8461           agctctgtag gattccttat gacgtattca ctgtctcagt cacgagaatc tttaaagcag
>NCBI old      #8461           agctctgtag gattccttat gacgtattca ctgtctcagt cacgagaatc tttaaagcag
                               ..........................................................

                #8461          agctctgtag gattccttat gacgtattca ctgtctcagt cacgagaatc tttaaagcag

>Part1 of Ens.#5761            taaatgcttt gcttttgtcc atcttgctat aatcaaaact gaccatgccc ggctaatttt
>NCBI b29      #8521           taaatgcttt gcttttgtcc atcttgctat aatcaaaact gaccatgccc ggctaatttt
>NCBI old      #8521           taaatgcttt gcttttgtcc atcttgctat aatcaaaact gaccatgccc ggctaatttt
                               ..........................................................

                #8521          taaatgcttt gcttttgtcc atcttgctat aatcaaaact gaccatgccc ggctaatttt

>Part1 of Ens.#5821            ttatttttttg tagagatccc tatgttgccc aggcttgtct cgagctcttg ggcacaaatg
>NCBI b29      #8581           ttatttttttg tagagatccc tatgttgccc aggcttgtct cgagctcttg ggcacaaatg
>NCBI old      #8581           ttatttttttg tagagatccc tatgttgccc aggcttgtct cgagctcttg ggcacaaatg
                               ..........................................................

                #8581          ttatttttttg tagagatccc tatgttgccc aggcttgtct cgagctcttg ggcacaaatg

>Part1 of Ens.#5881            atcttccccc cttgacctcc caaagttctg ggattacagg cgtgagccac tgtacccagc
>NCBI b29      #8641           atcttccccc cttgacctcc caaagttctg ggattacagg cgtgagccac tgtacccagc
>NCBI old      #8641           atcttccccc cttgacctcc caaagttctg ggattacagg cgtgagccac tgtacccagc
                               ..........................................................

                #8641          atcttccccc cttgacctcc caaagttctg ggattacagg cgtgagccac tgtacccagc

>Part1 of Ens.#5941            ccctttggat ttttaattgt tcaaactggt agtcatagtc ccacttcaga ggagagaacc
>NCBI b29      #8701           ccctttggat ttttaattgt tcaaactggt agtcatagtc ccacttcaga ggagagaacc
>NCBI old      #8701           ccctttggat ttttaattgt tcaaactggt agtcatagtc ccacttcaga ggagagaacc
                               ..........................................................

                #8701          ccctttggat ttttaattgt tcaaactggt agtcatagtc ccacttcaga ggagagaacc

>Part1 of Ens.#6001            atatatgtga ccagaattct tgggggaaaa gtaagaaact actcttaaaa gttccttaaa
>NCBI b29      #8761           atatatgtga ccagaattct tgggggaaaa gtaagaaact actcttaaaa gttccttaaa
>NCBI old      #8761           atatatgtga ccagaattct tgggggaaaa gtaagaaact actcttaaaa gttccttaaa
                               ..........................................................

                #8761          atatatgtga ccagaattct tgggggaaaa gtaagaaact actcttaaaa gttccttaaa

>Part1 of Ens.#6061            gggtagaatt ctcttttcta gaaagagtag cctagggttg tccaatcttt tggctttcgt
>NCBI b29      #8821           gggtagaatt ctcttttcta gaaagagtag cctagggttg tccaatcttt tggctttcgt
>NCBI old      #8821           gggtagaatt ctcttttcta gaaagagtag cctagggttg tccaatcttt tggctttcgt
                               ..........................................................
```

FIGURE 2J

```
                #8821       gggtagaatt ctctllltcta gaaagagtag cctagggttg tccaatcttt tggctttcgt

>Partl of Ens.#6121         gggccacact ggaaggagaa attatcttgg gccacacatt aaatatgcta acactaacaa
>NCBI b29      #8881        gggccacact ggaaggagaa attatcttgg gccacacatt aaatatgcta acactaacaa
>NCBI old      #8881        gggccacact ggaaggagaa attatcttgg gccacacatt aaatatgcta acactaacaa
                            ..........................................................
                #8881       gggccacact ggaaggagaa attatcttgg gccacacatt aaatatgcta acactaacaa

>Partl of Ens.#6181         tagctgatga gctaaaaaac aaacttttt tgatcacaca aaaaaaaatc tcataacgtt
>NCBI b29      #8941        tagctgatga gctaaaaaac aaacttttt tgatcacaca aaaaaaaatc tcataacgtt
>NCBI old      #8941        tagctgatga gctaaaaaac aaacttttt tgatcacaca aaaaaaaatc tcataacgtt
                            ..........................................................
                #8941       tagctgatga gctaaaaaac aaacttttt tgatcacaca aaaaaaaatc tcataacgtt

>Partl of Ens.#6241         ttaagaaagt ttacgaattt gtgttgggcc tcattcaaag ccatcctggg ccgcatgcgg
>NCBI b29      #9001        ttaagaaagt ttacgaattt gtgttgggcc tcattcaaag ccatcctggg ccgcatgcgg
>NCBI old      #9001        ttaagaaagt ttacgaattt gtgttgggcc tcattcaaag ccatcctggg ccgcatgcgg
                            ..........................................................
                #9001       ttaagaaagt ttacgaattt gtgttgggcc tcattcaaag ccatcctggg ccgcatgcgg

>Partl of Ens.#6301         cctgtgtgct atgggttaga caagcttaga gtagagcata gtgggtgaga gaagaggctt
>NCBI b29      #9061        cctgtgtgct atgggttaga caagcttaga gtagagcata gtgggtgaga gaagaggctt
>NCBI old      #9061        cctgtgtgct atgggttaga caagcttaga gtagagcata gtgggtgaga gaagaggctt
                            ..........................................................
                #9061       cctgtgtgct atgggttaga caagcttaga gtagagcata gtgggtgaga gaagaggctt

>Partl of Ens.#6361         cagatgcttc tggcttgagt tcaaatcctg actctgccac tcactagctg tgtgaccttt
>NCBI b29      #9121        cagatgcttc tggcttgagt tcaaatcctg actctgccac tcactagctg tgtgaccttt
>NCBI old      #9121        cagatgcttc tggcttgagt tcaaatcctg actctgccac tcactagctg tgtgaccttt
                            ..........................................................
                #9121       cagatgcttc tggcttgagt tcaaatcctg actctgccac tcactagctg tgtgaccttt

>Partl of Ens.#6421         ggcaagtgac ttaacctctc tgtgtttcca tttcctcctc tgtaaaatgg aaataataat
>NCBI b29      #9181        ggcaagtgac ttaacctctc tgtgtttcca tttcctcctc tgtaaaatgg aaataataat
>NCBI old      #9181        ggcaagtgac ttaacctctc tgtgtttcca tttcctcctc tgtaaaatgg aaataataat
                            ..........................................................
                #9181       ggcaagtgac ttaacctctc tgtgtttcca tttcctcctc tgtaaaatgg aaataataat

>Partl of Ens.#6481         aatgatacct gtctcataaa gtggtcagga ttaaatgagt taatacatat aaagagctta
>NCBI b29      #9241        aatgatacct gtctcataaa gtggtcagga ttaaatgagt taatacatat aaagagctta
>NCBI old      #9241        aatgatacct gtctcataaa gtggtcagga ttaaatgagt taatacatat aaagagctta
                            ..........................................................
                #9241       aatgatacct gtctcataaa gtggtcagga ttaaatgagt taatacatat aaagagctta

>Partl of Ens.#6541         gatcagtgcc tggtgcctat aaacactcca gtggtaactg ttataatttg tattatctct
>NCBI b29      #9301        gatcagtgcc tggtgcctat aaacactcca gtggtaactg ttataatttg tattatctct
>NCBI old      #9301        gatcagtgcc tggtgcctat aaacactcca gtggtaactg ttataatttg tattatctct
                            ..........................................................
                #9301       gatcagtgcc tggtgcctat aaacactcca gtggtaactg ttataatttg tattatctct

>Partl of Ens.#6601         tggtgattcc cataggtcac aggcattgaa agtgctcaaa gggcccaggtg cagtggctca
>NCBI b29      #9361        tggtgattcc cataggtcac aggcattgaa agtgctcaaa gggcccaggtg cagtggctca
>NCBI old      #9361        tggtgattcc cataggtcac aggcattgaa agtgctcaaa gggcccaggtg cagtggctca
                            ..........................................................
                #9361       tggtgattcc cataggtcac aggcattgaa agtgctcaaa gggcccaggtg cagtggctca

>Partl of Ens.#6661         cgcctgtaat cccagcactt tgggaggtcg aggtgggtgg atcacctgag gtcaggagtt
>NCBI b29      #9421        cgcctgtaat cccagcactt tgggaggtcg aggtgggtgg atcacctgag gtcaggagtt
>NCBI old      #9421        cgcctgtaat cccagcactt tgggaggtcg aggtgggtgg atcacctgag gtcaggagtt
                            ..........................................................
                #9421       cgcctgtaat cccagcactt tgggaggtcg aggtgggtgg atcacctgag gtcaggagtt

>Partl of Ens.#6721         cgagaccagc ctggccaaca tggtgaaacc ccgtctctac taaaaatacа aaattagcca
>NCBI b29      #9481        cgagaccagc ctggccaaca tggtgaaacc ccgtctctac taaaaatacа aaattagcca
>NCBI old      #9481        cgagaccagc ctggccaaca tggtgaaacc ccgtctctac taaaaatacа aaattagcca
                            ..........................................................
                #9481       cgagaccagc ctggccaaca tggtgaaacc ccgtctctac taaaaatacа aaattagcca

>Partl of Ens.#6781         ggcgtgttgg cgcatgcctg taatcccagc tacttgggaa gctgaggcag gagaattgcc
>NCBI b29      #9541        ggcgtgttgg cgcatgcctg taatcccagc tacttgggaa gctgaggcag gagaattgcc
>NCBI old      #9541        ggcgtgttgg cgcatgcctg taatcccagc tacttgggaa gctgaggcag gagaattgcc
                            ..........................................................
                #9541       ggcgtgttgg cgcatgcctg taatcccagc tacttgggaa gctgaggcag gagaattgcc

>Partl of Ens.#6841         tgaacccggg aggcggaggt tgcagtgagc caagatcaca ccattgcact ccagcttagg
>NCBI b29      #9601        tgaacccggg aggcggaggt tgcagtgagc caagatcaca ccattgcact ccagcttagg
>NCBI old      #9601        tgaacccggg aggcggaggt tgcagtgagc caagatcaca ccattgcact ccagcttagg
                            ..........................................................
                #9601       tgaacccggg aggcggaggt tgcagtgagc caagatcaca ccattgcact ccagcttagg

>Partl of Ens.#6901         caacaagagc gaaacatcgt ctcaaaaaaa aaaaaagaga aaatgctcaa atgtaataaa
>NCBI b29      #9661        caacaagagc gaaacatcgt ctcaaaaaaa aaaaaagaga aaatgctcaa atgtaataaa
>NCBI old      #9661        caacaagagc gaaacatcgt ctcaaaaaaa aaaaaagaga aaatgctcaa atgtaataaa
                            ..........................................................
```

FIGURE 2K

```
                #9661          caacaagagc gaaacatcgt ctcaaaaaaa aaaaaagaga aaatgctcaa atgtaataaa

>Part1 of Ens.#6961            gacagtattc atcgagcgtt ttctctgtgc catcagtgtg ctgagtgctt tacctgcacc
>NCBI b29      #9721            gacagtattc atcgagcgtt ttctctgtgc catcagtgtg ctgagtgctt tacctgcacc
>NCBI old      #9721            gacagtattc atcgagcgtt ttctctgtgc catcagtgtg ctgagtgctt tacctgcacc
                                ................................................................
                #9721            gacagtattc atcgagcgtt ttctctgtgc catcagtgtg ctgagtgctt tacctgcacc

>Part1 of Ens.#7021            atctcattta tttcccgcaa accctcctgt ggtttaggtg ctaccatccc catttggcag
>NCBI b29      #9781            atctcattta tttcccgcaa accctcctgt ggtttaggtg ctaccatccc catttggcag
>NCBI old      #9781            atctcattta tttcccgcaa accctcctgt ggtttaggtg ctaccatccc catttggcag
                                ................................................................
                #9781            atctcattta tttcccgcaa accctcctgt ggtttaggtg ctaccatccc catttggcag

>Part1 of Ens.#7081            aagtggaaat ggagtcccct agaaggtacg tagcttgtag aaagtcatgt aattagtcaa
>NCBI b29      #9841            aagtggaaat ggagtcccct agaaggtacg tagcttgtag aaagtcatgt aattagtcaa
>NCBI old      #9841            aagtggaaat ggagtcccct agaaggtacg tagcttgtag aaagtcatgt aattagtcaa
                                ................................................................
                #9841            aagtggaaat ggagtcccct agaaggtacg tagcttgtag aaagtcatgt aattagtcaa

>Part1 of Ens.#7141            catcgtactt ccagccactg gccagcagaa gcttccccat caattcaaat ccattgaatc
>NCBI b29      #9901            catcgtactt ccagccactg gccagcagaa gcttccccat caattcaaat ccattgaatc
>NCBI old      #9901            catcgtactt ccagccactg gccagcagaa gcttccccat caattcaaat ccattgaatc
                                ................................................................
                #9901            catcgtactt ccagccactg gccagcagaa gcttccccat caattcaaat ccattgaatc

>Part1 of Ens.#7201            ctaatggcaa tttaagaggc ctgttttatt ttataggtgg gtgtttgtgt gggaaaaggg
>NCBI b29      #9961            ctaatggcaa tttaagaggc ctgttttatt ttataggtgg gtgtttgtgt gggaaaaggg
>NCBI old      #9961            ctaatggcaa tttaagaggc ctgttttatt ttataggtgg gtgtttgtgt gggaaaaggg
                                ................................................................
                #9961            ctaatggcaa tttaagaggc ctgttttatt ttataggtgg gtgtttgtgt gggaaaaggg

>Part1 of Ens.#7261            ctaccaggaa actgactccg tggtcagctc cgttacgacc aaggtcaagg gcgtggctgt
>NCBI b29      #10021           ctaccaggaa actgactccg tggtcagctc cgttacgacc aaggtcaagg gcgtggctgt
>NCBI old      #10021           ctaccaggaa actgactccg tggtcagctc cgttacgacc aaggtcaagg gcgtggctgt
                                ................................................................
                #10021           ctaccaggaa actgactccg tggtcagctc cgttacgacc aaggtcaagg gcgtggctgt

>Part1 of Ens.#7321            gaccaacact tctaaacttg gattccggat ctgggatgtg gcggattatg tgataccagc
>NCBI b29      #10081           gaccaacact tctaaacttg gattccggat ctgggatgtg gcggattatg tgataccagc
>NCBI old      #10081           gaccaacact tctaaacttg gattccggat ctgggatgtg gcggattatg tgataccagc
                                ................................................................
                #10081           gaccaacact tctaaacttg gattccggat ctgggatgtg gcggattatg tgataccagc

>Part1 of Ens.#7381            tcaggtgtgt ctcccactgt gtcttctgtc taacactgac accttgctct ttactgactt
>NCBI b29      #10141           tcaggtgtgt ctcccactgt gtcttctgtc taacactgac accttgctct ttactgactt
>NCBI old      #10141           tcaggtgtgt ctcccactgt gtcttctgtc taacactgac accttgctct ttactgactt
                                ................................................................
                #10141           tcaggtgtgt ctcccactgt gtcttctgtc taacactgac accttgctct ttactgactt

>Part1 of Ens.#7441            tgcacacttg attaatgatt gacgtggcct gagtcctgac atcaccaatg cccagaatga
>NCBI b29      #10201           tgcacacttg attaatgatt gacgtggcct gagtcctgac atcaccaatg cccagaatga
>NCBI old      #10201           tgcacacttg attaatgatt gacgtggcct gagtcctgac atcaccaatg cccagaatga
                                ................................................................
                #10201           tgcacacttg attaatgatt gacgtggcct gagtcctgac atcaccaatg cccagaatga

>Part1 of Ens.#7501            ttgttttagt ccaagaaaac ggtgcagcca gatctctttt cagcttagga tctaacagag
>NCBI b29      #10261           ttgttttagt ccaagaaaac ggtgcagcca gatctctttt cagcttagga tctaacagag
>NCBI old      #10261           ttgttttagt ccaagaaaac ggtgcagcca gatctctttt cagcttagga tctaacagag
                                ................................................................
                #10261           ttgttttagt ccaagaaaac ggtgcagcca gatctctttt cagcttagga tctaacagag

>Part1 of Ens.#7561            tgtatccagg cagctgtggt tcctcagctt gaaagcatgg caggctgaaa gccctgaaag
>NCBI b29      #10321           tgtatccagg cagctgtggt tcctcagctt gaaagcatgg caggctgaaa gccctgaaag
>NCBI old      #10321           tgtatccagg cagctgtggt tcctcagctt gaaagcatgg caggctgaaa gccctgaaag
                                ................................................................
                #10321           tgtatccagg cagctgtggt tcctcagctt gaaagcatgg caggctgaaa gccctgaaag

>Part1 of Ens.#7621            atgagaaaac aaatgtctta aatcctgata gaaattcaag tggaaatatc accttccac
>NCBI b29      #10381           atgagaaaac aaatgtctta aatcctgata gaaattcaag tggaaatatc accttccac
>NCBI old      #10381           atgagaaaac aaatgtctta aatcctgata gaaattcaag tggaaatatc accttccac
                                ................................................................
                #10381           atgagaaaac aaatgtctta aatcctgata gaaattcaag tggaaatatc accttccac

>Part1 of Ens.#7681            attttgggtt tgcattattg tcccaaatac acgaggtaaa tcttttcaga cagtgtaaaa
>NCBI b29      #10441           attttgggtt tgcattattg tcccaaatac acgaggtaaa tcttttcaga cagtgtaaaa
>NCBI old      #10441           attttgggtt tgcattattg tcccaaatac acgaggtaaa tcttttcaga cagtgtaaaa
                                ................................................................
                #10441           attttgggtt tgcattattg tcccaaatac acgaggtaaa tcttttcaga cagtgtaaaa

>Part1 of Ens.#7741            actttaggcc gggcacagta gctcacgtgt gtaatcccag cactttggga ggctgggggcg
>NCBI b29      #10501           actttaggcc gggcacagta gctcacgtgt gtaatcccag cactttggga ggctgggggcg
>NCBI old      #10501           actttaggcc gggcacagta gctcacgtgt gtaatcccag cactttggga ggctgggggcg
                                ................................................................
```

FIGURE 2L

```
            #10501      actttaggcc gggcacagta gctcacgtgt gtaatcccag cactttggga ggctgggggg

>Part1 of Ens.#7801     ggcagatacc ttgagcccag gagtttgaga ccagcctggg caatgtggtg aaaccccatc
>NCBI b29     #10561     ggcagatacc ttgagcccag gagtttgaga ccagcctggg caatgtggtg aaaccccatc
>NCBI old     #10561     ggcagatacc ttgagcccag gagtttgaga ccagcctggg caatgtggtg aaaccccatc
                        ..........................................................
            #10561      ggcagatacc ttgagcccag gagtttgaga ccagcctggg caatgtggtg aaaccccatc

>Part1 of Ens.#7861     tctataaaaa aaaaaaaaaa gaaaaaaaaa gaaaagaatt agagaccatt agagaccaaa
>NCBI b29     #10621     tctataaaaa aaaaaaaaaa gaaaaaaaaa gaaaagaatt agagaccatt agagaccaaa
>NCBI old     #10621     tctataaaaa aaaaaaaaaa gaaaaaaaaa gaaaagaatt agagaccatt agagaccaaa
                        ..........................................................
            #10621      tctataaaaa aaaaaaaaaa gaaaaaaaaa gaaaagaatt agagaccatt agagaccaaa

>Part1 of Ens.#7921     acccaaacaa ggaggaaaca gagacagcag cacaggactc catcgggagc gaaaggaagg
>NCBI b29     #10681     acccaaacaa gaaggaaaca gagacagcag cacaggactc catcgggagc gaaaggaagg
>NCBI old     #10681     acccaaacaa gaaggaaaca gagacagcag cacaggactc catcgggagc gaaaggaagg
                        ..........................................................
            #10681      acccaaacaa gaaggaaaca gagacagcag cacaggactc catcgggagc gaaaggaagg

>Part1 of Ens.#7981     gaagacattg tttactgctg catactagtc ttcaggctgt taatctgcaa agctccaagt
>NCBI b29     #10741     gaagacattg tttactgctg catactagtc ttcaggctgt taatctgcaa agctccaagt
>NCBI old     #10741     gaagacattg tttactgctg catactagtc ttcaggctgt taatctgcaa agctccaagt
                        ..........................................................
            #10741      gaagacattg tttactgctg catactagtc ttcaggctgt taatctgcaa agctccaagt

>Part1 of Ens.#8041     taaacattga ctctttagtc ctggtaacag ctctgcacca cacagtcaaa cccgacatag
>NCBI b29     #10801     taaacattga ctctttagtc ctggtaacag ctctgcacca cacagtcaaa cccgacatag
>NCBI old     #10801     taaacattga ctctttagtc ctggtaacag ctctgcacca cacagtcaaa cccgacatag
                        ..........................................................
            #10801      taaacattga ctctttagtc ctggtaacag ctctgcacca cacagtcaaa cccgacatag

>Part1 of Ens.#8101     ccctccagat ggcctatgtg ctaagggacc gagggcgacc agtagagata agctgccaac
>NCBI b29     #10861     ccctccagat ggcctatgtg ctaagggacc gagggcgacc agtagagata agctgccaac
>NCBI old     #10861     ccctccagat ggcctatgtg ctaagggacc gagggcgacc agtagagata agctgccaac
                        ..........................................................
            #10861      ccctccagat ggcctatgtg ctaagggacc gagggcgacc agtagagata agctgccaac

>Part1 of Ens.#8161     tcttggctca gagctcttgc cagggaaaac aagaactcca gaaactcttt cttcaggtgt
>NCBI b29     #10921     tcttggctca gagctcttgc cagggaaaac aagaactcca gaaactcttt cttcaggtgt
>NCBI old     #10921     tcttggctca gagctcttgc cagggaaaac aagaactcca gaaactcttt cttcaggtgt
                        ..........................................................
            #10921      tcttggctca gagctcttgc cagggaaaac aagaactcca gaaactcttt cttcaggtgt

>Part1 of Ens.#8221     gttcttcccc ctgccttctc aagagaaggg tcttgtgggg tccccagaga ggagaccgag
>NCBI b29     #10981     gttcttcccc ctgccttctc aagagaaggg tcttgtgggg tccccagaga ggagaccgag
>NCBI old     #10981     gttcttcccc ctgccttctc aagagaaggg tcttgtgggg tccccagaga ggagaccgag
                        ..........................................................
            #10981      gttcttcccc ctgccttctc aagagaaggg tcttgtgggg tccccagaga ggagaccgag

>Part1 of Ens.#8281     ctaatgagcc ataaaagtca ggaggaaggg ccaggtgcgg tggctcatgc ctgtaatccc
>NCBI b29     #11041     ctaatgagcc ataaaagtca ggaggaaggg ccaggtgcgg tggctcatgc ctgtaatccc
>NCBI old     #11041     ctaatgagcc ataaaagtca ggaggaaggg ccaggtgcgg tggctcatgc ctgtaatccc
                        ..........................................................
            #11041      ctaatgagcc ataaaagtca ggaggaaggg ccaggtgcgg tggctcatgc ctgtaatccc

>Part1 of Ens.#8341     agcactttgg gaggccgagg tcggtgtatc atgaggtcaa gagattgaga ccatcctggc
>NCBI b29     #11101     agcactttgg gaggccgagg tcggtgtatc atgaggtcaa gagattgaga ccatcctggc
>NCBI old     #11101     agcactttgg gaggccgagg tcggtgtatc atgaggtcaa gagattgaga ccatcctggc
                        ..........................................................
            #11101      agcactttgg gaggccgagg tcggtgtatc atgaggtcaa gagattgaga ccatcctggc

>Part1 of Ens.#8401     taacatggtg aaaccctgtc tactaaaaat acaaaaatta gtggggtgtg gtggtgggtg
>NCBI b29     #11161     taacatggtg aaaccctgtc tactaaaaat acaaaaatta gtggggtgtg gtggtgggtg
>NCBI old     #11161     taacatggtg aaaccctgtc tactaaaaat acaaaaatta gtggggtgtg gtggtgggtg
                        ..........................................................
            #11161      taacatggtg aaaccctgtc tactaaaaat acaaaaatta gtggggtgtg gtggtgggtg

>Part1 of Ens.#8461     cctgtagtcc cagctactag ggaggctgag gcaggagaat ctcttgaacc tgggaggcgg
>NCBI b29     #11221     cctgtagtcc cagctactag ggaggctgag gcaggagaat ctcttgaacc tgggaggcgg
>NCBI old     #11221     cctgtagtcc cagctactag ggaggctgag gcaggagaat ctcttgaacc tgggaggcgg
                        ..........................................................
            #11221      cctgtagtcc cagctactag ggaggctgag gcaggagaat ctcttgaacc tgggaggcgg

>Part1 of Ens.#8521     aggttgtagt gagccgagat agtgccactg cactccagcc tggtgacaga gcaagactcc
>NCBI b29     #11281     aggttgtagt gagccgagat agtgccactg cactccagcc tggtgacaga gcaagactcc
>NCBI old     #11281     aggttgtagt gagccgagat agtgccactg cactccagcc tggtgacaga gcaagactcc
                        ..........................................................
            #11281      aggttgtagt gagccgagat agtgccactg cactccagcc tggtgacaga gcaagactcc

>Part1 of Ens.#8581     gtctcaaaaa aaaagatcaa gtggaacagg ggatttgcag ttggctggag ggggcagtcc
>NCBI b29     #11341     gtctcaaaaa aaaagatcaa gtggaacagg ggatttgcag ttggctggag ggggcagtcc
>NCBI old     #11341     gtctcaaaaa aaaagatcaa gtggaacagg ggatttgcag ttggctggag ggggcagtcc
                        ..........................................................
```

FIGURE 2M

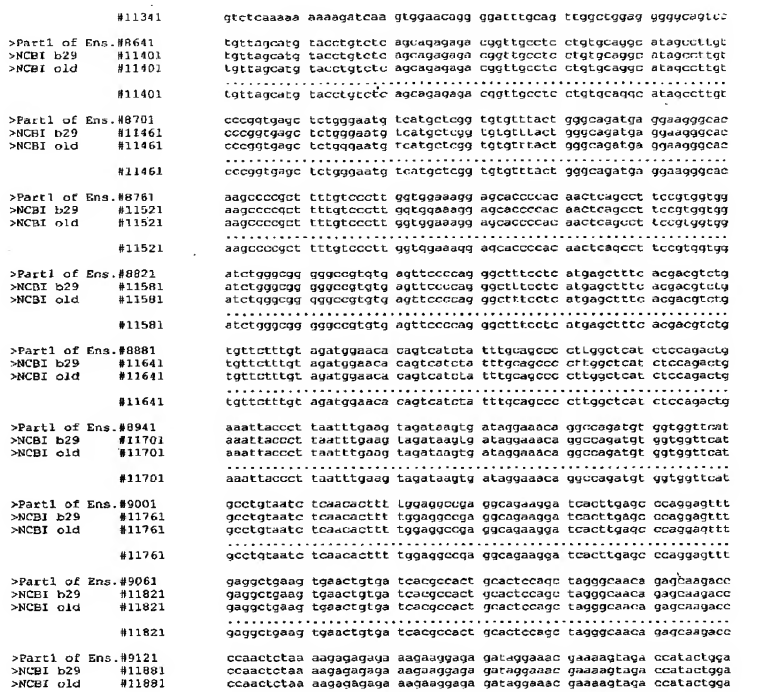


FIGURE 2N

```
            #12181      tgggcatgtc tacagactgg ttggtttctg agttgccctg atccccaaga tggctcccac

>Partl of Ens.#9481     tactgactgt tatcctcagg agtgtgaaat gctaaaagct ttgggcatct cctttaaaa
>NCBI b29     #12241     tactgactgt tatcctcagg agtgtgaaat gctaaaagct ttgggcatct cctttaaaa
>NCBI old     #12241     tactgactgt tatcctcagg agtgtgaaat gctaaaagct ttgggcatct cctttaaaa
                         ..........................................................
            #12241      tactgactgt tatcctcagg agtgtgaaat gctaaaagct ttgggcatct cctttaaaa

>Partl of Ens.#9541     tgcctgggcc cagccaggcg cagtggctta cgcctgtagt cccagcactt ttggaggcca
>NCBI b29     #12301     tgcctgggcc cagccaggcg cagtggctta cgcctgtagt cccagcactt ttggaggcca
>NCBI old     #12301     tgcctgggcc cagccaggcg cagtggctta cgcctgtagt cccagcactt ttggaggcca
                         ..........................................................
            #12301      tgcctgggcc cagccaggcg cagtggctta cgcctgtagt cccagcactt ttggaggcca

>Partl of Ens.#9601     aggcaggagg atcacttgag ctcaggattc aagaccagcc tgggcaatat agtgagaccc
>NCBI b29     #12361     aggcaggagg atcacttgag ctcaggattc aagaccagcc tgggcaatat agtgagaccc
>NCBI old     #12361     aggcaggagg atcacttgag ctcaggattc aagaccagcc tgggcaatat agtgagaccc
                         ..........................................................
            #12361      aggcaggagg atcacttgag ctcaggattc aagaccagcc tgggcaatat agtgagaccc

>Partl of Ens.#9661     catcttgatt ggatgggtgg tggatggatg gatggatgga tggatggatg gatggatgga
>NCBI b29     #12421     catcttgatt ggatgggtgg tggatggatg gatggatgga tggatggatg gatggatgga
>NCBI old     #12421     catcttgatt ggatgggtgg tggatggatg gatggatgga tggatggatg gatggatgga
                         ..........................................................
            #12421      catcttgatt ggatgggtgg tggatggatg gatggatgga tggatggatg gatggatgga

>Partl of Ens.#9721     tagatagata gatagataga tagatagata gatagataga tggaaagaaa gaaagaatag
>NCBI b29     #12481     tagatagata gatagataga tagatagata gatagataga tggaaagaaa gaaagaatag
>NCBI old     #12481     tagatagata gatagataga tagatagata gatagataga tggaaagaaa gaaagaatag
                         ..........................................................
            #12481      tagatagata gatagataga tagatagata gatagataga tggaaagaaa gaaagaatag

>Partl of Ens.#9781     ataaatgata ggataggaga aagaaaggga aagaaagaaa agagaagaga aaagaaagaa
>NCBI b29     #12541     ataaatgata ggataggaga aagaaaggga aagaaagaaa agagaagaga aaagaaagaa
>NCBI old     #12541     ataaatgata ggataggaga aagaaaggga aagaaagaaa agagaagaga aaagaaagaa
                         ..........................................................
            #12541      ataaatgata ggataggaga aagaaaggga aagaaagaaa agagaagaga aaagaaagaa

>Partl of Ens.#9841     aagggatagg atagatgaca gatgatagga tagattagat aggtgatага taatcgataa
>NCBI b29     #12601     aagggatagg atagatgaca gatgatagga tagattagat aggtgatага taatcgataa
>NCBI old     #12601     aagggatagg atagatgaca gatgatagga tagattagat aggtgatага taatcgataa
                         ..........................................................
            #12601      aagggatagg atagatgaca gatgatagga tagattagat aggtgatага taatcgataa

>Partl of Ens.#9901     atagaaaaca gatgatagat ataggataga ttagatagat agatagatag atagatagat
>NCBI b29     #12661     atagaaaaca gatgatagat ataggataga ttagatagat agatagatag atagatagat
>NCBI old     #12661     atagaaaaca gatgatagat ataggataga ttagatagat agatagatag atagatagat
                         ..........................................................
            #12661      atagaaaaca gatgatagat ataggataga ttagatagat agatagatag atagatagat

>Partl of Ens.#9961     agatagacag ataatagata gattagatta gatggatgga caccagcccc caatgagagg
>NCBI b29     #12721     agatagacag ataatagata gattagatta gatggatgga caccagcccc caatgagagg
>NCBI old     #12721     agatagacag ataatagata gattagatta gatggatgga caccagcccc caatgagagg
                         ..........................................................
            #12721      agatagacag ataatagata gattagatta gatggatgga caccagcccc caatgagagg

>Partl of Ens.#10021    aggatccaac tagaatcttt tggggtcttt gccttttatt ccttatatga ggaggtcatt
>NCBI b29     #12781     aggatccaac tagaatcttt tggggtcttt gccttttatt ccttatatga ggaggtcatt
>NCBI old     #12781     aggatccaac tagaatcttt tggggtcttt gccttttatt ccttatatga ggaggtcatt
                         ..........................................................
            #12781      aggatccaac tagaatcttt tggggtcttt gccttttatt ccttatatga ggaggtcatt

>Partl of Ens.#10081    atcagcaatt cttcagtctc atccctagca aattgctttc cagaaactct aaatcaatta
>NCBI b29     #12841     atcagcaatt cttcagtctc atccctagca aattgctttc cagaaactct aaatcaatta
>NCBI old     #12841     atcagcaatt cttcagtctc atccctagca aattgctttc cagaaactct aaatcaatta
                         ..........................................................
            #12841      atcagcaatt cttcagtctc atccctagca aattgctttc cagaaactct aaatcaatta

>Partl of Ens.#10141    tttggcacct ttatacatat aagcaggaaa cagaaccttt tcaaagagaa atgtcacttg
>NCBI b29     #12901     tttggcacct ttatacatat aaggaggaaa cagaaccttt tcaaagagaa atgtcacttg
>NCBI old     #12901     tttggcacct ttatacatat aaggaggaaa cagaaccttt tcaaagagaa atgtcacttg
                         ..........................................................
            #12901      tttggcacct ttatacatat aaggaggaaa cagaaccttt tcaaagagaa atgtcacttg

>Partl of Ens.#10201    ctgaccctga ccgagtcttc tctagaatgt ttgatttata atttcagtta ggaaagaaaa
>NCBI b29     #12961     ctgaccctga ccgagtcttc tctagaatgt ttgatttata atttcagtta ggaaagaaaa
>NCBI old     #12961     ctgaccctga ccgagtcttc tctagaatgt ttgatttata atttcagtta ggaaagaaaa
                         ..........................................................
            #12961      ctgaccctga ccgagtcttc tctagaatgt ttgatttata atttcagtta ggaaagaaaa

>Partl of Ens.#10261    ccagccagat ttgagcccca tgaggcctca tgttttatt cctctgagat aatctggttg
>NCBI b29     #13021     ccagccagat ttgagcccca tgaggcctca tgtttttatt cctctgagat aatctggttg
>NCBI old     #13021     ccagccagat ttgagcccca tgaggcctca tgtttttatt cctctgagat aatctggttg
                         ..........................................................
```

FIGURE 2O

```
                #13021      ccagccagat ttgagcccca tgaggcctca tgtttttatt cctctgagat aarctggttg

>Partl of Ens.#10321        gcagggaatg tttttagcac ttgttggctc tttccctcag taagtcaagt ttgattgaga
>NCBI b29      #13081        gcagggaatg tttttagcac ttgttggctc tttccctcag taagtcaagt ttgattgaga
>NCBI old      #13081        gcagggaatg tttttagcac ttgttggctc tttccctcag taagtcaagt ttgattgaga
                             ..........................................................
                #13081      gcagggaatg tttttagcac ttgttggctc tttccctcag taagtcaagt ttgattgaga

>Partl of Ens.#10381        aaccattgtt tcagtttgca gacccccccc ttttttttaac tgcagtcaaa ttcacttata
>NCBI b29      #13141        aaccattgtt tcagtttgca gacccccccc ttttttttaac tgcagtcaaa ttcacttata
>NCBI old      #13141        aaccattgtt tcagtttgca gacccccccc ttttttttaac tgcagtcaaa ttcacttata
                             ..........................................................
                #13141      aaccattgtt tcagtttgca gacccccccc ttttttttaac tgcagtcaaa ttcacttata

>Partl of Ens.#10441        gaaactaacc attttaagtc gggcacagtg gctcatgctt gtaatcccag cacttgggga
>NCBI b29      #13201        gaaactaacc attttaagtc gggcacagtg gctcatgctt gtaatcccag cacttgggga
>NCBI old      #13201        gaaactaacc attttaagtc gggcacagtg gctcatgctt gtaatcccag cacttgggga
                             ..........................................................
                #13201      gaaactaacc attttaagtc gggcacagtg gctcatgctt gtaatcccag cacttgggga

>Partl of Ens.#10501        ggctgaggca ggtagatcac ttgaggccag gagttggaga ccagcctagc caacatggtg
>NCBI b29      #13261        ggctgaggca ggtagatcac ttgaggccag gagttggaga ccagcctagc caacatggtg
>NCBI old      #13261        ggctgaggca ggtagatcac ttgaggccag gagttggaga ccagcctagc caacatggtg
                             ..........................................................
                #13261      ggctgaggca ggtagatcac ttgaggccag gagttggaga ccagcctagc caacatggtg

>Partl of Ens.#10561        aaatcctatc tctactaaag atacaaaaat tagccaggtg tggtgatgta cacctgtagt
>NCBI b29      #13321        aaatcctatc tctactaaag atacaaaaat tagccaggtg tggtgatgta cacctgtagt
>NCBI old      #13321        aaatcctatc tctactaaag atacaaaaat tagccaggtg tggtgatgta cacctgtagt
                             ..........................................................
                #13321      aaatcctatc tctactaaag atacaaaaat tagccaggtg tggtgatgta cacctgtagt

>Partl of Ens.#10621        cccagctact caggagggtg aggctggaga atcgcttgaa cccaggaggc ggagattgca
>NCBI b29      #13381        cccagctact caggagggtg aggctggaga atcgcttgaa cccaggaggc ggagattgca
>NCBI old      #13381        cccagctact caggagggtg aggctggaga atcgcttgaa cccaggaggc ggagattgca
                             ..........................................................
                #13381      cccagctact caggagggtg aggctggaga atcgcttgaa cccaggaggc ggagattgca

>Partl of Ens.#10681        gtaagccaag atcgtgccac tgccctgcag cctgggcgac agagcaaggc tctgtctcaa
>NCBI b29      #13441        gtaagccaag atcgtgccac tgccctgcag cctgggcgac agagcaaggc tctgtctcaa
>NCBI old      #13441        gtaagccaag atcgtgccac tgccctgcag cctgggcgac agagcaaggc tctgtctcaa
                             ..........................................................
                #13441      gtaagccaag atcgtgccac tgccctgcag cctgggcgac agagcaaggc tctgtctcaa

>Partl of Ens.#10741        ataaataaat aaataagaaa gaaattaacc attttaaaat gaaaaattca gcacatttag
>NCBI b29      #13501        ataaataaat aaataagaaa gaaattaacc attttaaaat gaaaaattca gcacatttag
>NCBI old      #13501        ataaataaat aaataagaaa gaaattaacc attttaaaat gaaaaattca gcacatttag
                             ..........................................................
                #13501      ataaataaat aaataagaaa gaaattaacc attttaaaat gaaaaattca gcacatttag

>Partl of Ens.#10801        tgcattcaca gtgttatgca agcatcacct ctatctagtt caaaacattt ccatcacccc
>NCBI b29      #13561        tgcattcaca gtgttatgca agcatcacct ctatctagtt caaaacattt ccatcacccc
>NCBI old      #13561        tgcattcaca gtgttatgca agcatcacct ctatctagtt caaaacattt ccatcacccc
                             ..........................................................
                #13561      tgcattcaca gtgttatgca agcatcacct ctatctagtt caaaacattt ccatcacccc

>Partl of Ens.#10861        aaagggaagc ctcgtatgtt acacaaccac tccccattcc ccacttctct cactccctgg
>NCBI b29      #13621        aaagggaagc ctcgtatgtt acacaaccac tccccattcc ccacttctct cactccctgg
>NCBI old      #13621        aaagggaagc ctcgtatgtt acacaaccac tccccattcc ccacttctct cactccctgg
                             ..........................................................
                #13621      aaagggaagc ctcgtatgtt acacaaccac tccccattcc ccacttctct cactccctgg

>Partl of Ens.#10921        caacccgtaa gtgctttgtc tctatggact tgactgttct ggaaatttca catcaatgta
>NCBI b29      #13681        caacccgtaa gtgctttgtc tctatggact tgactgttct ggaaatttca catcaatgta
>NCBI old      #13681        caacccgtaa gtgctttgtc tctatggact tgactgttct ggaaatttca catcaatgta
                             ..........................................................
                #13681      caacccgtaa gtgctttgtc tctatggact tgactgttct ggaaatttca catcaatgta

>Partl of Ens.#10981        atggtatcct gtgtggcctt tgtgtctggc ttctttcact cagcatcatg tttttggggt
>NCBI b29      #13741        atggtatcct gtgtggcctt tgtgtctggc ttctttcact cagcatcatg tttttggggt
>NCBI old      #13741        atggtatcct gtgtggcctt tgtgtctggc ttctttcact cagcatcatg tttttggggt
                             ..........................................................
                #13741      atggtatcct gtgtggcctt tgtgtctggc ttctttcact cagcatcatg tttttggggt

>Partl of Ens.#11041        tcatccatgt tgtagcctgt atcagtgctt tattcctttа ttttatttttt tgagatgaaa
>NCBI b29      #13801        tcatccatgt tgtagcctgt atcagtgctt tattcctttа ttttatttttt tgagatgaaa
>NCBI old      #13801        tcatccatgt tgtagcctgt atcagtgctt tattcctttа ttttatttttt tgagatgaaa
                             ..........................................................
                #13801      tcatccatgt tgtagcctgt atcagtgctt tattcctttа ttttatttttt tgagatgaaa

>Partl of Ens.#11101        tctcactctg tcacccaggc tggagtgcac tggcgtgatc tcggctcact gcaacctcca
>NCBI b29      #13861        tctcactctg tcacccaggc tggagtgcac tggcgtgatc tcggctcact gcaacctcca
>NCBI old      #13861        tctcactctg tcacccaggc tggagtgcac tggcgtgatc tcggctcact gcaacctcca
                             ..........................................................
```

FIGURE 2P

```
                #13861      tctcactctg tcacccaggc tggagtgcac tggcgtgatc tcggctcact gcaacctcca

>Part1 of Ens.#11161        cctcccaggt tccagtgatt ctcctggctc agccccctga gtagctggga ttacaggtgc
>NCBI b29      #13921        cctcccaggt tccagtgatt ctcctggctc agccccctga gtagctggga ttacaggtgc
>NCBI old      #13921        cctcccaggt tccagtgatt ctcctggctc agccccctga gtagctggga ttacaggtgc
                            ....................................................................
                #13921      cctcccaggt tccagtgatt ctcctggctc agccccctga gtagctggga ttacaggtgc

>Part1 of Ens.#11221        ccgccaccac acccggctcc ttttgtgtgt gcgtgtgtct gtgtgtgttt gtgtgtgtgt
>NCBI b29      #13981        ccgccaccac acccggctcc ttttgtgtgt gcgtgtgtct gtgtgtgttt gtgtgtgtgt
>NCBI old      #13981        ccgccaccac acccggctcc ttttgtgtgt gcgtgtgtct gtgtgtgttt gtgtgtgtgt
                            ....................................................................
                #13981      ccgccaccac acccggctcc ttttgtgtgt gcgtgtgtct gtgtgtgttt gtgtgtgtgt

>Part1 of Ens.#11281        gtgtgtgtgt gtgtgtgtgt ttttagtaca aagtttcacc atgttggcca ggctaatctg
>NCBI b29      #14041        gtgtgtgtgt gtgtgtgtgt ttttagtaca aagtttcacc atgttggcca ggctaatctg
>NCBI old      #14041        gtgtgtgtgt gtgtgtgtgt ttttagtaca aagtttcacc atgttggcca ggctaatctg
                            ....................................................................
                #14041      gtgtgtgtgt gtgtgtgtgt ttttagtaca aagtttcacc atgttggcca ggctaatctg

>Part1 of Ens.#11341        gaattcttga cctcaggtga tcctcctgcc tcagcctccc aaagtgctgg aattacaggt
>NCBI b29      #14101        gaattcttga cctcaggtga tcctcctgcc tcagcctccc aaagtgctgg aattacaggt
>NCBI old      #14101        gaattcttga cctcaggtga tcctcctgcc tcagcctccc aaagtgctgg aattacaggt
                            ....................................................................
                #14101      gaattcttga cctcaggtga tcctcctgcc tcagcctccc aaagtgctgg aattacaggt

>Part1 of Ens.#11401        gtgagccacc atgcctggac ccttcattcc tttttattca tggctgaatc acattccatc
>NCBI b29      #14161        gtgagccacc atgcctggac ccttcattcc tttttattca tggctgaatc acattccatc
>NCBI old      #14161        gtgagccacc atgcctggac ccttcattcc tttttattca tggctgaatc acattccatc
                            ....................................................................
                #14161      gtgagccacc atgcctggac ccttcattcc tttttattca tggctgaatc acattccatc

>Part1 of Ens.#11461        atgtgactag actaggacag acatttgggc tgtttccacc attttttttt ttttttgag
>NCBI b29      #14221        atgtgactag actaggacag acatttgggc tgtttccacc attttttttt ttttttgag
>NCBI old      #14221        atgtgactag actaggacag acatttgggc tgtttccacc attttttttt ttttttgag
                            ....................................................................
                #14221      atgtgactag actaggacag acatttgggc tgtttccacc attttttttt ttttttgag

>Part1 of Ens.#11521        acagagtctc gtgtcatcca ggctggagtg cagtggcgcg atgtcggctc gctgcaacct
>NCBI b29      #14281        acagagtctc gtgtcatcca ggctggagtg cagtggcgcg atgtcggctc gctgcaacct
>NCBI old      #14281        acagagtctc gtgtcatcca ggctggagty cagtggcgcg atgtcggctc gctgcaacct
                            ....................................................................
                #14281      acagagtctc gtgtcatcca ggctggagtg cagtggcgcg atgtcggctc gctgcaacct

>Part1 of Ens.#11581        ccgactcccg ggttcaagca attctcctgc ctcagcctcc caagtagctg ggactacagg
>NCBI b29      #14341        ccgactcccg ggttcaagca attctcctgc ctcagcctcc caagtagctg ggactacagg
>NCBI old      #14341        ccgactcccg ggttcaagca attctcctgc ctcagcctcc caagtagctg ggactacagg
                            ....................................................................
                #14341      ccgactcccg ggttcaagca attctcctgc ctcagcctcc caagtagctg ggactacagg

>Part1 of Ens.#11641        tgcccgccat cacacccggc taatttttgt atttttagta gagacggggg tttcactgtg
>NCBI b29      #14401        tgcccgccat cacacccggc taatttttgt atttttagta gagacggggg tttcactgtg
>NCBI old      #14401        tgcccgccat cacacccggc taatttttgt atttttagta gagacggggg tttcactgtg
                            ....................................................................
                #14401      tgcccgccat cacacccggc taatttttgt atttttagta gagacggggg tttcactgtg

>Part1 of Ens.#11701        ttgaccaggc tggtcttgaa ctcctgacct cgtgatccgc ccgcctcagc ctcccaaagt
>NCBI b29      #14461        ttgaccaggc tggtcttgaa ctcctgacct cgtgatccgc ccgcctcagc ctcccaaagt
>NCBI old      #14461        ttgaccaggc tggtcttgaa ctcctgacct cgtgatccgc ccgcctcagc ctcccaaagt
                            ....................................................................
                #14461      ttgaccaggc tggtcttgaa ctcctgacct cgtgatccgc ccgcctcagc ctcccaaagt

>Part1 of Ens.#11761        gctgggatta caggcttgag ccacccgcgc cgacctgttt ccacctttta gctcttgtga
>NCBI b29      #14521        gctgggatta caggcttgag ccacccgcgc cgacctgttt ccacctttta gctcttgtga
>NCBI old      #14521        gctgggatta caggcttgag ccacccgcgc cgacctgttt ccacctttta gctcttgtga
                            ....................................................................
                #14521      gctgggatta caggcttgag ccacccgcgc cgacctgttt ccacctttta gctcttgtga

>Part1 of Ens.#11821        aaagtgttgc tctgaatatt tatgtacaag gacttgtttg agtcactgtt ttccattggg
>NCBI b29      #14581        aaagtgttgc tctgaatatt tatgtacaag gacttgtttg agtcactgtt ttccattggg
>NCBI old      #14581        aaagtgttgc tctgaatatt tatgtacaag gacttgtttg agtcactgtt ttccattggg
                            ....................................................................
                #14581      aaagtgttgc tctgaatatt tatgtacaag gacttgtttg agtcactgtt ttccattggg

>Part1 of Ens.#11881        ttataaactt tgatgggaga gagcgggaga gcacgcggtc agtgtttgag ttgtcctctt
>NCBI b29      #14641        ttataaactt tgatgggaga gagcgggaga gcacgcggtc agtgtttgag ttgtcctctt
>NCBI old      #14641        ttataaactt tgatgggaga gagcgggaga gcacgcggtc agtgtttgag ttgtcctctt
                            ....................................................................
                #14641      ttataaactt tgatgggaga gagcgggaga gcacgcggtc agtgtttgag ttgtcctctt

>Part1 of Ens.#11941        cagtcagcgt ctgcctttct tggctctccg tgagtcctct gagcgtggct tgcccgtgct
>NCBI b29      #14701        cagtcagcgt ctgcctttct tggctctccg tgagtcctct gagcgtggct tgcccgtgct
>NCBI old      #14701        cagtcagcgt ctgcctttct tggctctccg tgagtcctct gagcgtggct tgcccgtgct
                            ....................................................................
```

FIGURE 2Q

```
                #14701      cagtcagcgt ctgccrttct tggctctccg tgagtcctct gagcgtggct tgcccgtgct

>Partl of Ens.#12001        gtctcccctc tgcaggagga aaactcccctc ttcgtcatga ccaacgtgat cctcaccatg
>NCBI b29      #14761        gtctcccctc tgcaggagga aaactcccctc ttcgtcatga ccaacgtgat cctcaccatg
>NCBI old      #14761        gtctcccctc tgcaggagga aaactcccctc ttcgtcatga ccaacgtgat cctcaccatg
                            ..........................................................
                #14761      gtctcccctc tgcaggagga aaactcccctc ttcgtcatga ccaacgtgat cctcaccatg

>Partl of Ens.#12061        aaccagacac agggcctgtg ccccgaggta ggaggccccc gggaagagcc ccaggccccc
>NCBI b29      #14821        aaccagacac agggcctgtg ccccgaggta ggaggccccc gggaagagcc ccaggccccc
>NCBI old      #14821        aaccagacac agggcctgtg ccccgaggta ggaggccccc gggaagagcc ccaggccccc
                            ..........................................................
                #14821      aaccagacac agggcctgtg ccccgaggta ggaggccccc gggaagagcc ccaggccccc

>Partl of Ens.#12121        caccccctcc cacgcctgtc cacctgtgtg tgggccggg ccacgtggac tttcttttcg
>NCBI b29      #14881        caccccctctc cacgcctgtc cacctgtgtg tgggccgggg ccacgtggac tttcttttcg
>NCBI old      #14881        caccccctctc cacgcctgtc cacctgtgtg tgggccgggg ccacgtggac tttcttttcg
                            ..........................................................
                #14881      caccccctctc cacgcctgtc cacctgtgtg tgggccgggg ccacgtggac tttcttttcg

>Partl of Ens.#12181        ctccttcttt ttccagattc cagatgccgac cactgtgtgt aaatcagatg ccagctgtac
>NCBI b29      #14941        ctccttcttt ttccagattc cagatgccgac cactgtgtgt aaatcagatg ccagctgtac
>NCBI old      #14941        ctccttcttt ttccagattc cagatgccgac cactgtgtgt aaatcagatg ccagctgtac
                            ..........................................................
                #14941      ctccttcttt ttccagattc cagatgccgac cactgtgtgt aaatcagatg ccagctgtac

>Partl of Ens.#12241        tgccggctct gccggcaccc acagcaacgg tacgagcttg tggcctcctg gggagggcgg
>NCBI b29      #15001        tgccggctct gccggcaccc acagcaacgg tacgagcttg tggcctcctg gggagggcgg
>NCBI old      #15001        tgccggctct gccggcaccc acagcaacgg tacgagcttg tggcctcctg gggagggcgg
                            ..........................................................
                #15001      tgccggctct gccggcaccc acagcaacgg tacgagcttg tggcctcctg gggagggcgg

>Partl of Ens.#12301        cccctgagca gatcgccccc actgtggagc gtctctgata gagaaatctt cccaattcct
>NCBI b29      #15061        cccctgagca gatcgccccc actgtggagc gtctctgata gagaaatctt cccaattcct
>NCBI old      #15061        cccctgagca gatcgccccc actgtggagc gtctctgata gagaaatctt cccaattcct
                            ..........................................................
                #15061      cccctgagca gatcgccccc actgtggagc gtctctgata gagaaatctt cccaattcct

>Partl of Ens.#12361        tcacatgacc ctgggtgagc caggtgccga ggctgaggtc ctggagcccc tctacattca
>NCBI b29      #15121        tcacatgacc ctgggtgagc caggtgccga ggctgaggtc ctggagcccc tctacattca
>NCBI old      #15121        tcacatgacc ctgggtgagc caggtgccga ggctgaggtc ctggagcccc tctacattca
                            ..........................................................
                #15121      tcacatgacc ctgggtgagc caggtgccga ggctgaggtc ctggagcccc tctacattca

>Partl of Ens.#12421        ctgctgtcat tggagcccca caagccatcc cagctcttgc cctactgttt ttttttttgt
>NCBI b29      #15181        ctgctgtcat tggagcccca caagccatcc cagctcttgc cctactgttt ttttttttgt
>NCBI old      #15181        ctgctgtcat tggagcccca caagccatcc cagctcttgc cctactgttt ttttttttgt
                            ..........................................................
                #15181      ctgctgtcat tggagcccca caagccatcc cagctcttgc cctactgttt ttttttttgt

>Partl of Ens.#12481        tgttgttgtt gttttttctt gttttttttt ttgttttgtt tttttgtttt tttgagccag
>NCBI b29      #15241        tgttgttgtt gttttttctt gttttttttt ttgttttgtt tttttgtttt tttgagccag
>NCBI old      #15241        tgttgttgtt gttttttctt gttttttttt ttgttttgtt tttttgtttt tttgagccag
                            ..........................................................
                #15241      tgttgttgtt gttttttctt gttttttttt ttgttttgtt tttttgtttt tttgagccag

>Partl of Ens.#12541        agtctcgctc tgtcacccaa actggagtgc aatggcatga tctcggctca ctgcaacctc
>NCBI b29      #15301        agtctcgctc tgtcacccaa actggagtgc aatggcatga tctcggctca ctgcaacctc
>NCBI old      #15301        agtctcgctc tgtcacccaa actggagtgc aatggcatga tctcggctca ctgcaacctc
                            ..........................................................
                #15301      agtctcgctc tgtcacccaa actggagtgc aatggcatga tctcggctca ctgcaacctc

>Partl of Ens.#12601        cacctcccag tttcaaacaa ttctcctgcc tcagcctccc gagtagctgg gattacaggc
>NCBI b29      #15361        cacctcccag tttcaaacaa ttctcctgcc tcagcctccc gagtagctgg gattacaggc
>NCBI old      #15361        cacctcccag tttcaaacaa ttctcctgcc tcagcctccc gagtagctgg gattacaggc
                            ..........................................................
                #15361      cacctcccag tttcaaacaa ttctcctgcc tcagcctccc gagtagctgg gattacaggc

>Partl of Ens.#12661        atgcaccacc acgcccagct aatttttgta ctttggtag agatgggttt taccatgttg
>NCBI b29      #15421        atgcaccacc acgcccagct aatttttgta ctttggtag agatgggttt taccatgttg
>NCBI old      #15421        atgcaccacc acgcccagct aatttttgta ctttggtag agatgggttt taccatgttg
                            ..........................................................
                #15421      atgcaccacc acgcccagct aatttttgta ctttggtag agatgggttt taccatgttg

>Partl of Ens.#12721        gccaggctgg tctcgaactc ctgacctcag gtgatccacc cgcctcagcc tcccaaagtg
>NCBI b29      #15481        gccaggctgg tctcgaactc ctgacctcag gtgatccacc cgcctcagcc tcccaaagtg
>NCBI old      #15481        gccaggctgg tctcgaactc ctgacctcag gtgatccacc cgcctcagcc tcccaaagtg
                            ..........................................................
                #15481      gccaggctgg tctcgaactc ctgacctcag gtgatccacc cgcctcagcc tcccaaagtg

>Partl of Ens.#12781        ctaggattac aggcatgacc cactgtgcct ggcctcttgt cccattcttt agcttggcat
>NCBI b29      #15541        ctaggattac aggcatgacc cactgtgcct ggcctcttgt cccattcttt agcttggcat
>NCBI old      #15541        ctaggattac aggcatgacc cactgtgcct ggcctcttgt cccattcttt agcttggcat
                            ..........................................................
```

FIGURE 2R

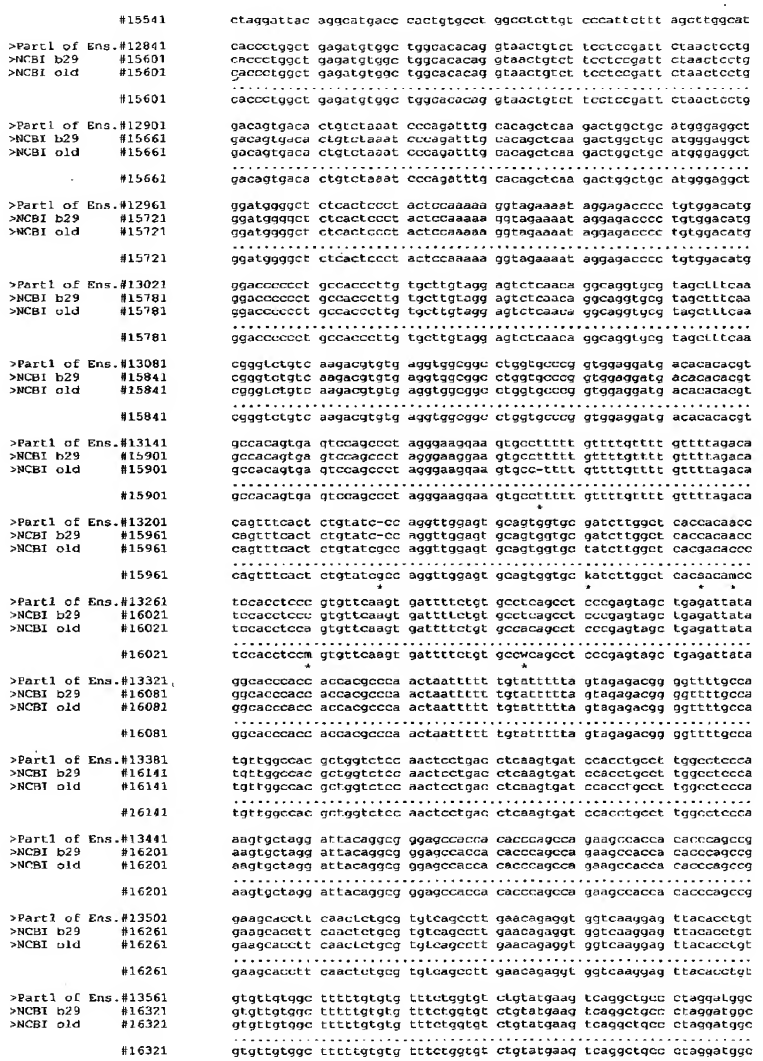


FIGURE 2S

```
                #16301      tgctctcaac cacaaggcag cgaccatcag tggcaagatg ggccacaagt agtctgtcac

>Part1 of Ens.#13601        tcacgagaag ggtgcaggtg actgacttcc ttgtatgtta ggactctgtc agttgcaaat
>NCBI b29       #16441      tcacgagaag ggtgcaggtg actgacttcc ttgtatgtta ggactctgtc agttgcaaat
>NCBI old       #16441      tcacgagaag ggtgcaggtg actgacttcc ttgtatgtta ggactctgtc agttgcaaat
                            ..........................................................
                #16441      tcacgagaag ggtgcaggtg actgacttcc ttgtatgtta ggactctgtc agttgcaaat

>Part1 of Ens.#13741        aactgacagg aaacagttct cagtgcctta agcacatgtg cacaaacaca tacacacata
>NCBI b29       #16501      aactgacagg aaacagttct cagtgcctta agcacatgtg cacaaacaca tacacacata
>NCBI old       #16501      aactgacagg aaacagttct cagtgcctta agcacatgtg cacaaacaca tacacacata
                            ..........................................................
                #16501      aactgacagg aaacagttct cagtgcctta agcacatgtg cacaaacaca tacacacata

>Part1 of Ens.#13801        cacatgcgtg cacagcgggg gggcagtgta ttgtccggct gaccaaaaag atccagcgtg
>NCBI b29       #16561      cacatgcgtg cacagcgggg gggcagtgta ttgtccggct gaccaaaaag atccagcgtg
>NCBI old       #16561      cacatgcgtg cacagcgggg gggcagtgta ttgtccggct gaccaaaaag atccagcgtg
                            ..........................................................
                #16561      cacatgcgtg cacagcgggg gggcagtgta ttgtccggct gaccaaaaag atccagcgtg

>Part1 of Ens.#13861        gccaggagaa tctcgtgcca ggctgaggca gaagaatctc ttgaacctgg gaggaggagg
>NCBI b29       #16621      gccaggagaa tctcgtgcca ggctgaggca gaagaatctc ttgaacctgg gaggaggagg
>NCBI old       #16621      gccaggagaa tctcgtgcca ggctgaggca gaagaatctc ttgaacctgg gaggaggagg
                            ..........................................................
                #16621      gccaggagaa tctcgtgcca ggctgaggca gaagaatctc ttgaacctgg gaggaggagg

>Part1 of Ens.#13921        ttgctgagct gagatggtgc cactgcactc cctcctgggc gacagaggcc actgcactcc
>NCBI b29       #16681      ttgctgagct gagatggtgc cactgcactc cctcctgggc gacagaggcc actgcactcc
>NCBI old       #16681      ttgctgagct gagatggtgc cactgcactc cctcctgggc gacagaggcc actgcactcc
                            ..........................................................
                #16681      ttgctgagct gagatggtgc cactgcactc cctcctgggc gacagaggcc actgcactcc

>Part1 of Ens.#13981        tgcctgggcg acagagtgag agacaaagga caacacccta gtctcccttc ctggctctgc
>NCBI b29       #16741      tgcctgggcg acagagtgag agacaaagga caacacccta gtctcccttc ctggctctgc
>NCBI old       #16741      tgcctgggcg acagagtgag agacaaagga caacacccta gtctcccttc ctggctctgc
                            ..........................................................
                #16741      tgcctgggcg acagagtgag agacaaagga caacacccta gtctcccttc ctggctctgc

>Part1 of Ens.#14041        atcccgccgt gccagcttca ctcgcaggac ccacagcata gcccgggagg cagcgctgca
>NCBI b29       #16801      atcccgccgt gccagcttca ctcgcaggac ccacagcata gcccgggagg cagcgctgca
>NCBI old       #16801      atcccgccgt gccagcttca ctcgcaggac ccacagcata gcccgggagg cagcgctgca
                            ..........................................................
                #16801      atcccgccgt gccagcttca ctcgcaggac ccacagcata gcccgggagg cagcgctgca

>Part1 of Ens.#14101        cttacagccc cccaagtctg agtctccctt tccagctgtc cctggagatt ctagacctgt
>NCBI b29       #16861      cttacagccc cccaagtctg agtctccctt tccagctgtc cctggagatt ctagacctgt
>NCBI old       #16861      cttacagccc cccaagtctg agtctccctt tccagctgtc cctggagatt ctagacctgt
                            ..........................................................
                #16861      cttacagccc cccaagtctg agtctccctt tccagctgtc cctggagatt ctagacctgt

>Part1 of Ens.#14161        ccatcatcgt atcaatggcc atatctctca acccatctct gtggctgggg aatgggtttt
>NCBI b29       #16921      ccatcatcgt atcaatggcc atatctctca acccatctct gtggctgggg aatgggtttt
>NCBI old       #16921      ccatcatcgt atcaatggcc atatctctca acccatctct gtggctgggg aatgggtttt
                            ..........................................................
                #16921      ccatcatcgt atcaatggcc atatctctca acccatctct gtggctgggg aatgggtttt

>Part1 of Ens.#14221        gcttatcagc ttgggcctgg gtcacatgcc cacctctgca gctggggcta gagtcagacc
>NCBI b29       #16981      gcttatcagc ttgggcctgg gtcacatgcc cacctctgca gctggggcta gagtcagacc
>NCBI old       #16981      gcttatcagc ttgggcctgg gtcacatgcc cacctctgca gctggggcta gagtcagacc
                            ..........................................................
                #16981      gcttatcagc ttgggcctgg gtcacatgcc cacctctgca gctggggcta gagtcagacc

>Part1 of Ens.#14281        cacccaagtg acataggcaa gcattgggga gatatggctt ccaaaggaca tttgggattc
>NCBI b29       #17041      cacccaagtg acataggcaa gcattgggga gatatggctt ccaaaggaca tttgggattc
>NCBI old       #17041      cacccaagtg acataggcaa gcattgggga gatatggctt ccaaaggaca tttgggattc
                            ..........................................................
                #17041      cacccaagtg acataggcaa gcattgggga gatatggctt ccaaaggaca tttgggattc

>Part1 of Ens.#14341        tgttaccaga aaagctggga gcagatgctg ggaagaaaac aacggatgtt tgctacagtg
>NCBI b29       #17101      tgttaccaga aaagctggga gcagatgctg ggaagaaaac aacggatgtt tgctacagtg
>NCBI old       #17101      tgttaccaga aaagctggga gcagatgctg ggaagaaaac aacggatgtt tgctacagtg
                            ..........................................................
                #17101      tgttaccaga aaagctggga gcagatgctg ggaagaaaac aacggatgtt tgctacagtg

>Part1 of Ens.#14401        tattttaaac taaactagac aggttcatcc tggcgcggtg gctcatgctg taatcctaac
>NCBI b29       #17161      tattttaaac taaactagac aggttcatcc tggcgcggtg gctcatgctg taatcctaac
>NCBI old       #17161      tattttaaac taaactagac aggttcatcc tggcgcggtg gctcatgctg taatcctaac
                            ..........................................................
                #17161      tattttaaac taaactagac aggttcatcc tggcgcggtg gctcatgctg taatcctaac

>Part1 of Ens.#14461        actttgggag gctgaggcag tcggatcatc tgaggccagg agttcaagac cagcctggcc
>NCBI b29       #17221      actttgggag gctgaggcag tcggatcatc tgaggccagg agttcaagac cagcctggcc
>NCBI old       #17221      actttgggag gctgaggcag tcggatcatc tgaggccagg agttcaagac cagcctggcc
                            ..........................................................
```

FIGURE 2T

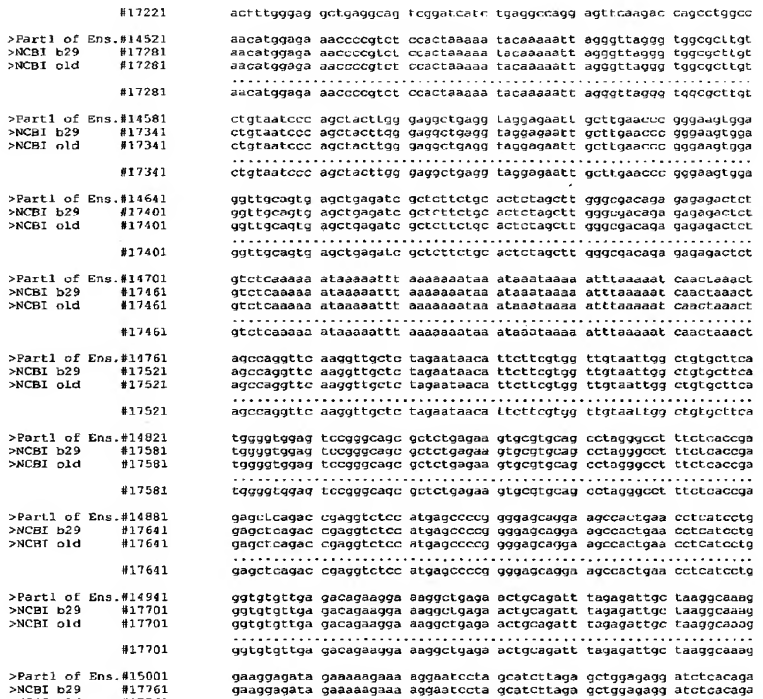


FIGURE 2U

```
                #18061      tgagtagctg ggattatagg cacccccccac cacaccttgc taatttttgt atttttagta

>Part1 of Ens.#15361        gagatgcagg tttcaccatg ttggccaggc tggtctcgaa ctcctgacct caggttatcc
>NCBI b29      #18121        gagatgcagg tttcaccatg ttggccaggc tggtctcgaa ctcctgacct caggttatcc
>NCBI old      #18121        gagatgcagg tttcaccatg ttggccaggc tggtctcgaa ctcctgacct caggttatcc
                            ..........................................................................
                #18121      gagatgcagg tttcaccatg ttggccaggc tggtctcgaa ctcctgacct caggttatcc

>Part1 of Ens.#15421        acctgcctca gcctcccaaa gtgctgggat tataggtgtg agccactgca cccagcccga
>NCBI b29      #18181        acctgcctca gcctcccaaa gtgctgggat tataggtgtg agccactgca cccagcccga
>NCBI old      #18181        acctgcctca gcctcccaaa gtgctgggat tataggtgtg agccactgca cccagcccga
                            ..........................................................................
                #18181      acctgcctca gcctcccaaa gtgctgggat tataggtgtg agccactgca cccagcccga

>Part1 of Ens.#15481        ttctgtttat tatttattta ttttttattt atttgtattt atttatttat ttgagatgga
>NCBI b29      #18241        ttctgtttat tatttattta ttttttattt atttgtattt atttatttat ttgagatgga
>NCBI old      #18241        ttctgtttat tatttattta ttttttattt atttgtattt atttatttat ttgagatgga
                            ..........................................................................
                #18241      ttctgtttat tatttattta ttttttattt atttgtattt atttatttat ttgagatgga

>Part1 of Ens.#15541        gtctcgctct gtcgcccagg ctggagtgca gtggcgccat ctctgttcac tgcaagctcc
>NCBI b29      #18301        gtctcgctct gtcgcccagg ctggagtgca gtggcgccat ctctgttcac tgcaagctcc
>NCBI old      #18301        gtctcgctct gtcgcccagg ctggagtgca gtggcgccat ctctgttcac tgcaagctcc
                            ..........................................................................
                #18301      gtctcgctct gtcgcccagg ctggagtgca gtggcgccat ctctgttcac tgcaagctcc

>Part1 of Ens.#15601        gcctcctgcg ctcacgccat tctcctgcct cagcctcccc agtagctggg actacaggcg
>NCBI b29      #18361        gcctcctgcg ctcacgccat tctcctgcct cagcctcccc agtagctggg actacaggcg
>NCBI old      #18361        gcctcctgcg ctcacgccat tctcctgcct cagcctcccc agtagctggg actacaggcg
                            ..........................................................................
                #18361      gcctcctgcg ctcacgccat tctcctgcct cagcctcccc agtagctggg actacaggcg

>Part1 of Ens.#15661        cccaccacca cacccagcta atttttttgt atttttagta gagaggggqt ttcactgtgt
>NCBI b29      #18421        cccaccacca cacccagcta atttttttgt atttttagta gagaggggqt ttcactgtgt
>NCBI old      #18421        cccaccacca cacccagcta atttttttgt atttttagta gagaggggqt ttcactgtgt
                            ..........................................................................
                #18421      cccaccacca cacccagcta atttttttgt atttttagta gagaggggqt ttcactgtgt

>Part1 of Ens.#15721        tagccaggat ggtctcgatc tcctgacctc gtgatccacc tgccttggcc tcccaaagtg
>NCBI b29      #18481        tagccaggat ggtctcgatc tcctgacctc gtgatccacc tgccttggcc tcccaaagtg
>NCBI old      #18481        tagccaggat ggtctcgatc tcctgacctc gtgatccacc tgccttggcc tcccaaagtg
                            ..........................................................................
                #18481      tagccaggat ggtctcgatc tcctgacctc gtgatccacc tgccttggcc tcccaaagtg

>Part1 of Ens.#15781        ctgggattgc acgcgtgagt caccacgccc ggccctgaat atttattatg aaaaaaagaa
>NCBI b29      #18541        ctgggattgc acgcgtgagt caccacgccc ggccctgaat atttattatg aaaaaaagaa
>NCBI old      #18541        ctgggattgc acgcgtgagt caccacgccc ggccctgaat atttattatg aaaaaaagaa
                            ..........................................................................
                #18541      ctgggattgc acgcgtgagt caccacgccc ggccctgaat atttattatg aaaaaaagaa

>Part1 of Ens.#15841        tgtataatat ctcattgatt tttgtatgtt ggttacatgt tgaagtaata ttttgggtta
>NCBI b29      #18601        tgtataatat ctcattgatt tttgtatgtt ggttacatgt tgaagtaata ttttgggtta
>NCBI old      #18601        tgtataatat ctcattgatt tttgtatgtt ggttacatgt tgaagtaata ttttgggtta
                            ..........................................................................
                #18601      tgtataatat ctcattgatt tttgtatgtt ggttacatgt tgaagtaata ttttgggtta

>Part1 of Ens.#15901        ctgtataaaa ttttacttgt ttcttttact ttgtagtgtg gctattaqqa aactttaagt
>NCBI b29      #18661        ctgtataaaa ttttacttgt ttcttttact ttgtagtgtg gctattaqqa aactttaagt
>NCBI old      #18661        ctgtataaaa ttttacttgt ttcttttact ttgtagtgtg gctattaqqa aactttaagt
                            ..........................................................................
                #18661      ctgtataaaa ttttacttgt ttcttttact ttgtagtgtg gctattaqqa aactttaagt

>Part1 of Ens.#15961        tggagtgtgg cttgtagttg tgtctcataa gatacttttt ttttgacaca ggatcttgct
>NCBI b29      #18721        tggagtgtgg cttgtagttg tgtctcataa gatacttttt ttttgacaca ggatcttgct
>NCBI old      #18721        tggagtgtgg cttgtagttg tgtctcataa gatacttttt ttttgacaca ggatcttgct
                            ..........................................................................
                #18721      tggagtgtgg cttgtagttg tgtctcataa gatacttttt ttttgacaca ggatcttgct

>Part1 of Ens.#16021        ctgtcaccca gtgcagtgat aacgatcatg gctcactgca gcttcgaact cctggactca
>NCBI b29      #18781        ctgtcaccca gtgcagtgat aacgatcatg gctcactgca gcttcgaact cctggactca
>NCBI old      #18781        ctgtcaccca gtgcagtgat aacgatcatg gctcactgca gcttcgaact cctggactca
                            ..........................................................................
                #18781      ctgtcaccca gtgcagtgat aacgatcatg gctcactgca gcttcgaact cctggactca

>Part1 of Ens.#16081        agtgatcttc cqgcctcatc ctcctgtgta qctgggacca cagggqcgca ccaccatgcc
>NCBI b29      #18841        agtgatcttc ccgcctcatc ctcctgtgta gctgggacca cagggqcgca ccaccatgcc
>NCBI old      #18841        agtgatcttc ccgcctcatc ctcctgtgta gctgggacca cagggqcgca ccaccatgcc
                            ..........................................................................
                #18841      agtgatcttc ccgcctcatc ctcctgtgta gctgggacca cagggqcgca ccaccatgcc

>Part1 of Ens.#16141        cagccaactt atttttgta gagacagggt ctcaccatgt tgcccaggct ggtcttgaac
>NCBI b29      #18901        cagccaactt atttttgta gagacagggt ctcaccatgt tgcccaggct ggtcttgaac
>NCBI old      #18901        cagccaactt atttttgta gagacagggt ctcaccatgt tgcccaggct ggtcttgaac
                            ..........................................................................
```

FIGURE 2V

```
                  #18901     cagccaactt atttttttgta gagacagggt ctcaccatgt tgcccaggct ggtcttgaac

>Part1 of Ens.#16201     tcctgggctc aagcaaccca cctgccttag cctcccaaag tgctgggatt acagcgtgag
>NCBI b29      #18961     tcctgggctc aagcaaccca cctgccttag cctcccaaag tgctgggatt acagcgtgag
>NCBI old      #18961     tcctgggctc aagcaaccca cctgccttag cctcccaaag tgctgggatt acagcgtgag
                          ............................................................
               #18961     tcctgggctc aagcaaccca cctgccttag cctcccaaag tgctgggatt acagcgtgag

>Part1 of Ens.#16261     ccaccacacc tggcctcaca atatactttt attggacatc actgttgcag atctaatgcg
>NCBI b29      #19021     ccaccacacc tggcctcaca atatactttt attggacatc actgttgcag atctaatgcg
>NCBI old      #19021     ccaccacacc tggcctcaca atatactttt attggacatc actgttgcag atctaatgcg
                          ............................................................
               #19021     ccaccacacc tggcctcaca atatactttt attggacatc actgttgcag atctaatgcg

>Part1 of Ens.#16321     gtgtaaaatg atttttaagt gcacactgat ataagaaata acattggcca ggcacag
>Part2 of Ens.>#1>                                                                   tgg
>NCBI b29      #19081     gtgtaaaatg atttttaagt gcacactgat ataagaaata acattggcca ggcacagtgg
>NCBI old      #19081     gtgtaaaatg atttttaagt gcacactgat ataagaaata acattggcca ggcacagtgg
                          ............................................................
               #19081     gtgtaaaatg atttttaagt gcacactgat ataagaaata acattggcca ggcacagtgg

>Part2 of Ens.#4         ctcatg-ctg taatcctaac actttgggag gctgaggcag tcggatcatt tgaggccagg
>NCBI b29      #19141     ctcatg-ctg taatcctaac actttgggag gctgaggcag tcggatcatt tgaggccagg
>NCBI old      #19141     ctcatgccta taatcccagc actttgggag gcagagacgg gtggatcact tgaggccagg
                          ............................................................
               #19141     ctcatgcctr taatccyarc actttgggag gcwgagrcrg kyggatcayt tgaggccagg
                                *    *          * *            *    * *  **     *
>Part2 of Ens.#64        agttcaaatt cagcctggcc aacatggtga aacccctct ctactaaaaa tataaaaatt
>NCBI b29      #19201     agttcaaatt cagcctggcc aacatggtga aacccctct ctactaaaaa tataaaaatt
>NCBI old      #19201     agttcaaaac cagcctggcc aacatggtga aacccctct ctactaaaaa tataaaaatt
                          ............................................................
               #19201     agttcaaawy cagcctggcc aacatggtga aacccctct ctactaaaaa tataaaaatt
                                   **
>Part2 of Ens.#124       agccaggcgt ggtggtgcgc acctgtaatc ccactactca ggaggctgag gcacgagaat
>NCBI b29      #19261     agccaggcgt ggtggtgcgc acctgtaatc ccactactca ggaggctgag gcacgagaat
>NCBI old      #19261     agccaggcgt ggtggtgcgc acctgtaatc ccactactca ggaggctgag gcacgagaat
                          ............................................................
               #19261     agccaggcgt ggtggtgcgc acctgtaatc ccactactca ggaggctgag gcacgagaat

>Part2 of Ens.#184       tgcttgaacc cgggagatag aagttgcagt gagccaagat catgccactg cgctccagcc
>NCBI b29      #19321     tgcttgaacc cgggagatag aagttgcagt gagccaagat catgccactg cgctccagcc
>NCBI old      #19321     tgcttgaacc cgggagatag aagttgcagt gagccaagat catgccactg cgctccagcc
                          ............................................................
               #19321     tgcttgaacc cgggagatag aagttgcagt gagccaagat catgccactg cgctccagcc

>Part2 of Ens.#244       cgggtgacag aataagaccc tgtctcaaaa acagaacaaa gactaaaacg ttttggaact
>NCBI b29      #19381     cgggtgacag aataagaccc tgtctcaaaa acagaacaaa gactaaaacg ttttggaact
>NCBI old      #19381     cgggtgacag aataagaccc tgtctcaaaa acagaacaaa gactaaaacg ttttggaact
                          ............................................................
               #19381     cgggtgacag aataagaccc tgtctcaaaa acagaacaaa gactaaaacg ttttggaact

>Part2 of Ens.#304       agacagaggt gatgattgca caacattatg aatgtactaa atgccacaga attgtccagt
>NCBI b29      #19441     agacagaggt gatgattgca caacattatg aatgtactaa atgccacaga attgtccagt
>NCBI old      #19441     agacagaggt gatgattgca caacattatg aatgtactaa atgccacaga attgtccagt
                          ............................................................
               #19441     agacagaggt gatgattgca caacattatg aatgtactaa atgccacaga attgtccagt

>Part2 of Ens.#364       ttaaaatggt taattttatg ttatgtgaat ttcatctcaa tttctctaaa aaggtgagtt
>NCBI b29      #19501     ttaaaatggt taattttatg ttatgtgaat ttcatctcaa tttctctaaa aaggtgagtt
>NCBI old      #19501     ttaaaatggt taattttatg ttatgtgaat ttcatctcaa tttctctaaa aaggtgagtt
                          ............................................................
               #19501     ttaaaatggt taattttatg ttatgtgaat ttcatctcaa tttctctaaa aaggtgagtt

>Part2 of Ens.#424       gatcttaaag tattacatgt gaggtacaca ggtgtgtgtg tgggtgttac ggtacaggaa
>NCBI b29      #19561     gatcttaaag tattacatgt gaggtacaca ggtgtgtgtg tgggtgttac ggtacaggaa
>NCBI old      #19561     gatcttaaag tattacatgt gaggtacaca ggtgtgtgtg tgggtgttac ggtacaggaa
                          ............................................................
               #19561     gatcttaaag tattacatgt gaggtacaca ggtgtgtgtg tgggtgttac ggtacaggaa

>Part2 of Ens.#484       atgacaaaaa cttgaacacc cctgttgtaa gccatcccctc ttgagggagt ggggactttg
>NCBI b29      #19621     atgacaaaaa cttgaacacc cctgttgtaa gccatcccctc ttgagggagt ggggactttg
>NCBI old      #19621     atgacaaaaa cttgaacacc cctgttgtaa gccatcccctc ttgagggagt ggggactttg
                          ............................................................
               #19621     atgacaaaaa cttgaacacc cctgttgtaa gccatcccctc ttgagggagt ggggactttg

>Part2 of Ens.#544       aagacctgag agaagtccag gcacagtggc ccatgcctgt aatcccagca ctttgggaga
>NCBI b29      #19681     aagacctgag agaagtccag gcacagtggc ccatgcctgt aatcccagca ctttgggaga
>NCBI old      #19681     aagacctgag agaagtccag gcacagtggc ccatgcctgt aatcccagca ctttgggaga
                          ............................................................
               #19681     aagacctgag agaagtccag gcacagtggc ccatgcctgt aatcccagca ctttgggaga

>Part2 of Ens.#604       ccgaggtggg tggatcatct gaggtcagga gtttgagacc agcctgacca acatggtgaa
>NCBI b29      #19741     ccgaggtggg tggatcatct gaggtcagga gtttgagacc agcctgacca acatggtgaa
>NCBI old      #19741     ccgaggtggg tggatcatct gaggtcagga gtttgagacc agcctgacca acatggtgaa
```

FIGURE 2W

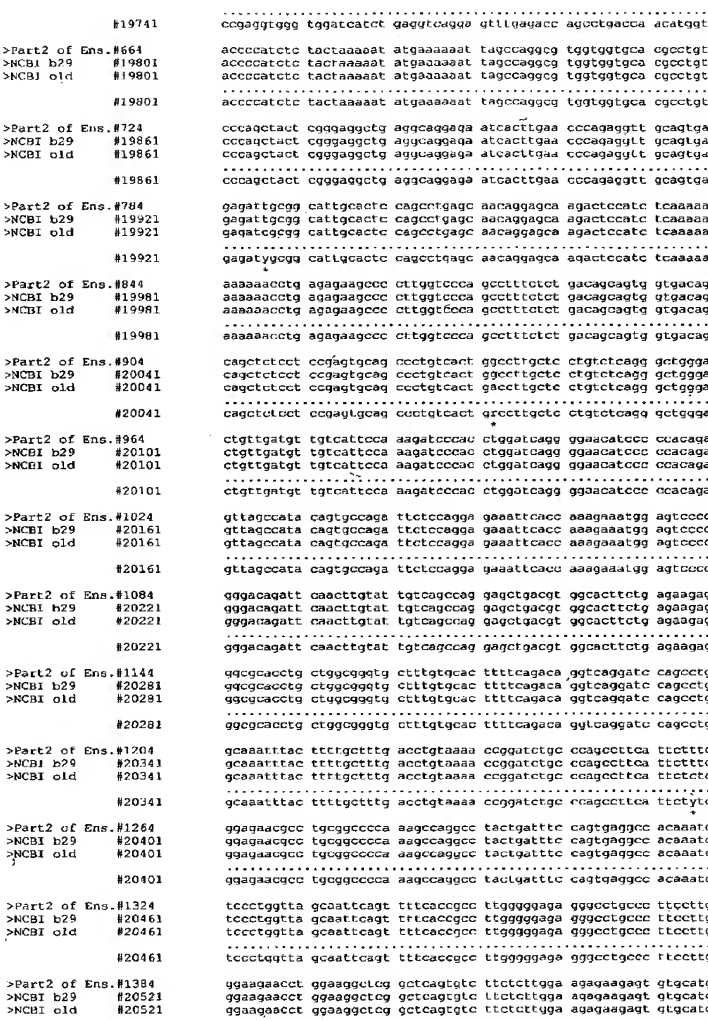


FIGURE 2X

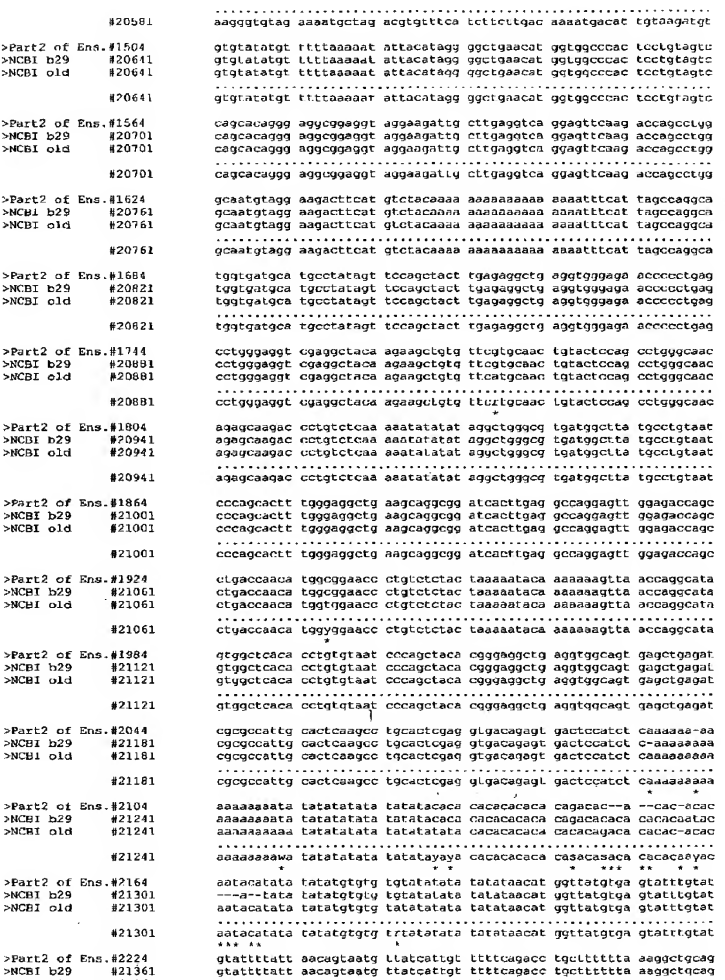


FIGURE 2Y

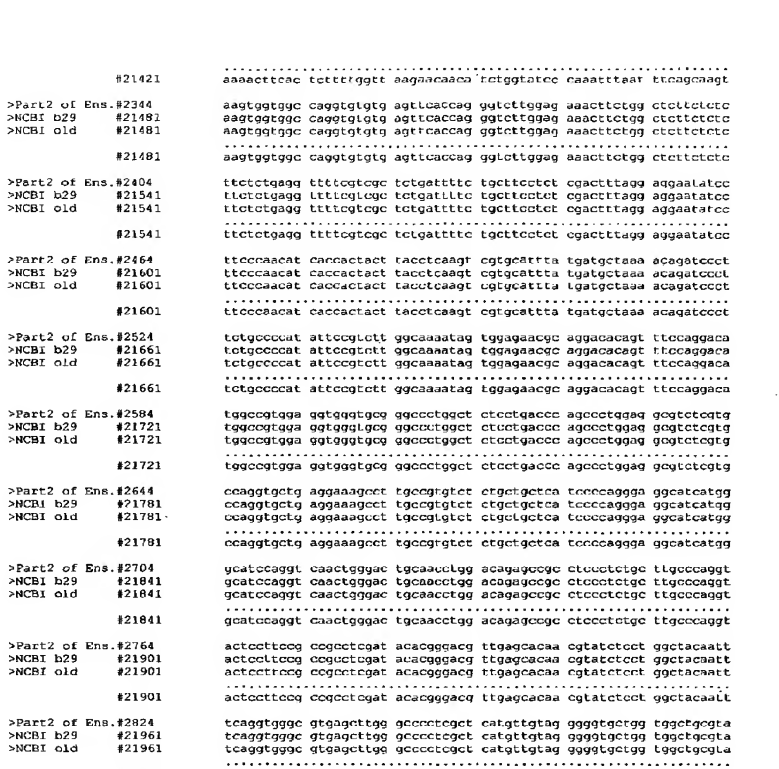


FIGURE 2Z

```
                                 ............................................................
                 #22261          ctcagccatg acctccattc actgcctcaa ggagcggatg atcttgtgat cctccagtcc

>Part2 of Ens.#3184              aaaggcctct ggggcctggc ccaggaattg gtttctcaaa ggttgaacct gtgccagaat
>NCBI b29        #22321          aaaggcctct ggggcctggc ccaggaattg gtttctcaaa ggttgaacct gtgccagaat
>NCBI old        #22321          aaaggcctct ggggcctggc ccaggaattg gtttctcaaa ggttgaacct gtgccagaat
                                 ............................................................
                 #22321          aaaggcctct ggggcctggc ccaggaattg gtttctcaaa ggttgaacct gtgccagaat

>Part2 of Ens.#3244              ctccagagtg caggggacac aggtctgagg cccctgaaag gccttgctgg gttccctgct
>NCBI b29        #22381          ctccagagtg caggggacac aggtctgagg cccctgaaag gccttgctgg gttccctgct
>NCBI old        #22381          ctccagagtg caggggacac aggtctgagg cccctgaaag gccttgctgg gttccctgct
                                 ............................................................
                 #22381          ctccagagtg caggggacac aggtctgagg cccctgaaag gccttgctgg gttccctgct

>Part2 of Ens.#3304              gcaaatgctg gcatctcggt gtcctggggc agatgtaagc gagtcccaca aacagtggct
>NCBI b29        #22441          gcaaatgctg gcatctcggt gtcctggggc agatgtaagc gagtcccaca aacagtggct
>NCBI old        #22441          gcaaatgctg gcatctcggt gtcctggggc agatgtaagc gagtcccaca aacagtggct
                                 ............................................................
                 #22441          gcaaatgctg gcatctcggt gtcctggggc agatgtaagc gagtcccaca aacagtggct

>Part2 of Ens.#3364              tgcaactaca ggacttgatt ctctcacagt tctcacagtt ctagaggcca caacctaaaa
>NCBI b29        #22501          tgcaactaca ggacttgatt ctctcacagt tctcacagtt ctagaggcca caacctaaaa
>NCBI old        #22501          tgcaactaca ggacttgatt ctctcacagt tctcacagtt ctagaggcca caacctaaaa
                                 ............................................................
                 #22501          tgcaactaca ggacttgatt ctctcacagt tctcacagtt ctagaggcca caacctaaaa

>Part2 of Ens.#3424              tccaggcatc agcagggctg tgctcgctct gaaggctctg gagagagtcc ttacttgtct
>NCBI b29        #22561          tccaggcatc agcagggctg tgctcgctct gaaggctctg gagagagtcc ttacttgtct
>NCBI old        #22561          tccaggcatc agcagggctg tgctcgctct gaaggctctg gagagagtcc ttacttgtct
                                 ............................................................
                 #22561          tccaggcatc agcagggctg tgctcgctct gaaggctctg gagagagtcc ttacttgtct

>Part2 of Ens.#3484              ctcctggctt ctggtagccc acggtgttcc ttggcttgtg gatgcatcac tccagtctct
>NCBI b29        #22621          ctcctggctt ctggtagccc acggtgttcc ttggcttgtg gatgcatcac tccagtctct
>NCBI old        #22621          ctcctggctt ctggtagccc acggtgttcc ttggcttgtg gatgcatcac tccagtctct
                                 ............................................................
                 #22621          ctcctggctt ctggtagccc acggtgttcc ttggcttgtg gatgcatcac tccagtctct

>Part2 of Ens.#3544              gcctctcctg tcatgaggcc ttctgccttg tacgtgtctc tgtgtctttt cttcttgtaa
>NCBI b29        #22681          gcctctcctg tcatgaggcc ttctgccttg tacgtgtctc tgtgtctttt cttcttgtaa
>NCBI old        #22681          gcctctcctg tcatgaggcc ttctgccttg tatgtgtctc tgtgtctttt cttcttgtaa
                                 ............................................................
                 #22681          gcctctcctg tcatgaggcc ttctgccttg taygtgtctc tgtgtctttt cttcttgtaa
                                                                   *
>Part2 of Ens.#3604              ggatatcagt cactgaattt agggcccacc ctactctagt atgacctcat cttaacaaat
>NCBI b29        #22741          ggatatcagt cactgaattt agggcccacc ctactctagt atgacctcat cttaacaaat
>NCBI old        #22741          ggatatcagt cactgaattt agggcccacc ctactctagt atgacctcat cttaacaaat
                                 ............................................................
                 #22741          ggatatcagt cactgaattt agggcccacc ctactctagt atgacctcat cttaacaaat

>Part2 of Ens.#3664              tacgtctgcc aatatcccat ttccaaataa ggtcacactc ccaagcttca ggtagatgtg
>NCBI b29        #22801          tacgtctgcc aatatcccat ttccaaataa ggtcacactc ccaagcttca ggtagatgtg
>NCBI old        #22801          tacgtctgcc aatatcccat ttccaaataa ggtcacactc ccaagcttca ggtagatgtg
                                 ............................................................
                 #22801          tacgtctgcc aatatcccat ttccaaataa ggtcacactc ccaagcttca ggtagatgtg

>Part2 of Ens.#3724              aatttttggg gacacactgt tcaacccact acagctgatg atctcatgat cagaatctat
>NCBI b29        #22861          aatttttggg gacacactgt tcaacccact acagctgatg atctcatgat cagaatctat
>NCBI old        #22861          aatttttggg gacacactgt tcaacccact acagctgatg atctcatgat cagaatctat
                                 ............................................................
                 #22861          aatttttggg gacacactgt tcaacccact acagctgatg atctcatgat cagaatctat

>Part2 of Ens.#3784              agccaggtaa atttccacgt accctttgtt gcatttttca cagtccttaa aaaggaagcg
>NCBI b29        #22921          agccaggtaa atttccacgt accctttgtt gcatttttca cagtccttaa aaaggaagcg
>NCBI old        #22921          agccaggtaa atttccacgt accctttgtt gcatttttca cagtccttaa aaaggaagcg
                                 ............................................................
                 #22921          agccaggtaa atttccacgt accctttgtt gcatttttca cagtccttaa aaaggaagcg

>Part2 of Ens.#3844              ctccaggccg ggcgcagtga ctcacacctg taatcctagc actttgggag gccaaggcgg
>NCBI b29        #22981          ctccaggccg ggcgcagtga ctcacacctg taatcctagc actttgggag gccaaggcgg
>NCBI old        #22981          ctccaggccg ggcgcagtga ctcacacctg taatcctagc actttgggag gccaaggcgg
                                 ............................................................
                 #22981          ctccaggccg ggcgcagtga ctcacacctg taatcctagc actttgggag gccaaggcgg

>Part2 of Ens.#3904              gcagatcacc tgaggccagc atggcgaaac cccatctcta ctaaaaaaaa tacaaaaatt
>NCBI b29        #23041          gcagatcacc tgaggccagc atggcgaaac cccatctcta ctaaaaaaaa tacaaaaatt
>NCBI old        #23041          gcagatcacc tgaggccagc atggcgaaac cccatctcta ctaaaaaaaa tacaaaaatt
                                 ............................................................
                 #23041          gcagatcacc tgaggccagc atggcgaaac cccatctcta ctaaaaaaaa tacaaaaatt

>Part2 of Ens.#3964              agctgggcat ggcggcgcat gcctgtaatc ccagcaactt gggaggctga ggcaggagaa
>NCBI b29        #23101          agctgggcat ggcggcgcat gcctgtaatc ccagcaactt gggaggctga ggcaggagaa
>NCBI old        #23101          agctgggcat ggtggcgcat gcctgtaatc ccagcaactt gggaggctga ggcaggagaa
```

FIGURE 2AA

```
                                                                                   ........................
                 #23101      agctgggcat ggyggcgcat gcctgtaatc ccagcaactt gggaggctga ggcaggagaa
                                          *

>Part2 of Ens.#4024          tcgcttgagc cagggaggca gaggttgcag tgagccgaga tcgcactgct gcactccagc
>NCBI b29     #23161         tcgcttgagc cagggaggca gaggttgcag tgagccgaga tcgcactgct gcactccagc
>NCBI old     #23161         tcgcttgagc cagggaggca gaggttgcag tgagccgaga tcgcactgct gcactccagc
                             ............................................................
              #23161         tcgcttgagc cagggaggca gaggttgcag tgagccgaga tcgcactgct gcactccagc

>Part2 of Ens.#4084          ctgggcgaca aagcaagtct ccgtctcagg gggaaaaaaa agaggccctg caaggaccat
>NCBI b29     #23221         ctgggcgaca aagcaagtct ccgtctcagg gggaaaaaaa agaggccctg caaggaccat
>NCBI old     #23221         ctgggcgaca aagcaagtct ccgtctcagg gggaaaaaaa agaggccctg caaggaccat
                             ............................................................
              #23221         ctgggcgaca aagcaagtct ccgtctcagg gggaaaaaaa agaggccctg caaggaccat

>Part2 of Ens.#4144          ccctgggcgc ggtcagcggt gctttgggct gcacttgcct ccccctagtg gtgacgtgtg
>NCBI b29     #23281         ccctgggcgc ggtcagcggt gctttgggct gcacttgcct ccccctagtg gtgacgtgtg
>NCBI old     #23281         ccctgggcgc ggtcagcggt gctttgggct gcacttgcct ccccctagtg gtgacgtgtg
                             ............................................................
              #23281         ccctgggcgc ggtcagcggt gctttgggct gcacttgcct ccccctagtg gtgacgtgtg

>Part2 of Ens.#4204          gaatccgatc catacgggat cctaagacct ctgtaattgt gggatctgct gtcctttga
>NCBI b29     #23341         gaatccgatc catacgggat cctaagacct ctgtaattgt gggatctgct gtcctttga
>NCBI old     #23341         gaatccgatc catacgggat cctaagacct ctgtaattgt gggatctgct gtcctttga
                             ............................................................
              #23341         gaatccgatc catacgggat cctaagacct ctgtaattgt gggatctgct gtcctttga

>Part2 of Ens.#4264          accttctggc tcactgagta aagcagatat ttgaattttt cttttctagc agttagattt
>NCBI b29     #23401         accttctggc tcactgagta aagcagatat ttgaattttt cttttctagc agttagattt
>NCBI old     #23401         accttctggc tcactgagta aagcagatat ttgaattttt cttttctagc agttagattt
                             ............................................................
              #23401         accttctggc tcactgagta aagcagatat ttgaattttt cttttctagc agttagattt

>Part2 of Ens.#4324          gtgttattcc atctctactg tcacacactt cttgtggaaa cctcaagctt aggtcagagc
>NCBI b29     #23461         gtgttattcc atctctactg tcacacactt cttgtggaaa cctcaagctt aggtcagagc
>NCBI old     #23461         gtgttattcc atctctactg tcacacactt cttgtggaaa cctcaagctt aggtcagagc
                             ............................................................
              #23461         gtgttattcc atctctactg tcacacactt cttgtggaaa cctcaagctt aggtcagagc

>Part2 of Ens.#4384          cctaaaacgg aaagacctca ggacgcctat cccccaacct cgggcagcac caggccctgc
>NCBI b29     #23521         cctaaaacgg aaagacctca ggacgcctat cccccaacct cgggcagcac caggccctgc
>NCBI old     #23521         cctaaaacgg aaagaactca ggacgcctat cccccaacct cgggcagcac caggccctgc
                             ............................................................
              #23521         cctaaaacgg aaagamctca ggacgcctat cccccaacct cgggcagcac caggccctgc
                                             *

>Part2 of Ens.#4444          ctgtgtctcc caggcctggc ccatggcgtg gaagttgtgg gccagataga gcatggagct
>NCBI b29     #23581         ctgtgtctcc caggcctggc ccatggcgtg gaagttgtgg gccagataga gcatggagct
>NCBI old     #23581         ctgtgtctcc caggcctggc ccatggcgtg gaagttgtgg gccagataga gcatggagct
                             ............................................................
              #23581         ctgtgtctcc caggcctggc ccatggcgtg gaagttgtgg gccagataga gcatggagct

>Part2 of Ens.#4504          agggtgcggg ccgaggctcc tgatggatcc ctggagaaca tggggcagca gatggtatgt
>NCBI b29     #23641         agggtgcggg ccgaggctcc tgatggatcc ctggagaaca tggggcagca gatggtatgt
>NCBI old     #23641         agggtgcggg ccgaggctcc tgatggatcc ctggagaaca tggggcagca gatggtatgt
                             ............................................................
              #23641         agggtgcggg ccgaggctcc tgatggatcc ctggagaaca tggggcagca gatggtatgt

>Part2 of Ens.#4564          cttgacggtt cttcaccaac catccatccc ctgctcatgc gtgtgtgctc tctcgcgtgc
>NCBI b29     #23701         cttgacggtt cttcaccaac catccatccc ctgctcatgc gtgtgtgctc tctcgcgtgc
>NCBI old     #23701         cttgacggtt cttcaccaac catccatccc ctgctcatgc gtgtgcgctc tctcgcgtgc
                             ............................................................
              #23701         cttgacggtt cttcaccaac catccatccc ctgctcatgc gtgtgygctc tctcgcgtgc
                                                                              *

>Part2 of Ens.#4624          tcgctctctc tctctctcat atatgtatat gtatatatgt gtatatatat atatatatag
>NCBI b29     #23761         tcgctctctc tctctctcat atatgtatat gtatatatgt gtatatatat atatatatag
>NCBI old     #23761         tcgctctctc tctctctcat atatgtatat gtatatatgt g--t--atat atatatatag
                             ............................................................
              #23761         tcgctctctc tctctctcat atatgtatat gtatatatgt gtatatatat atatatatag
                                                                        ** **

>Part2 of Ens.#4684          ccattatata tatagccatt atatatatat atattttttt ttcttcttct tttttaaggg
>NCBI b29     #23821         ccattatata tatagccatt atatatatat atattttttt ttcttcttct tttttaaggg
>NCBI old     #23821         ccattatata tatagccatt atatatatat at-ttttttt tttttcttct tttttaaggg
                             ............................................................
              #23821         ccattatata tatagccatt atatatatat atattttttt ttyttcttct tttttaaggg
                                                            *          *

>Part2 of Ens.#4744          agtgacttgc tctgtcaccc aggttggagt gcagtggtgc gatttcagct cactgcagcc
>NCBI b29     #23881         agtgacttgc tctgtcaccc aggttggagt gcagtggtgc gatttcagct cactgcagcc
>NCBI old     #23881         agtgacttgc tctgtcaccc aggttggagt gcagtggtgc gatttcagct cactgcagcc
                             ............................................................
              #23881         agtgacttgc tctgtcaccc aggttggagt gcagtggtgc gatttcagct cactgcagcc

>Part2 of Ens.#4804          tttgcctccc gggttcaaga cattctcctg cctcagccgc ctgagtagct gggattacag
>NCBI b29     #23941         tttgcctccc gggttcaaga cattctcctg cctcagccgc ctgagtagct gggattacag
>NCBI old     #23941         tttgcctccc gggttcaaga cattctcctg cctcagccgc ctgagtagct gggattacag
```

FIGURE 2BB

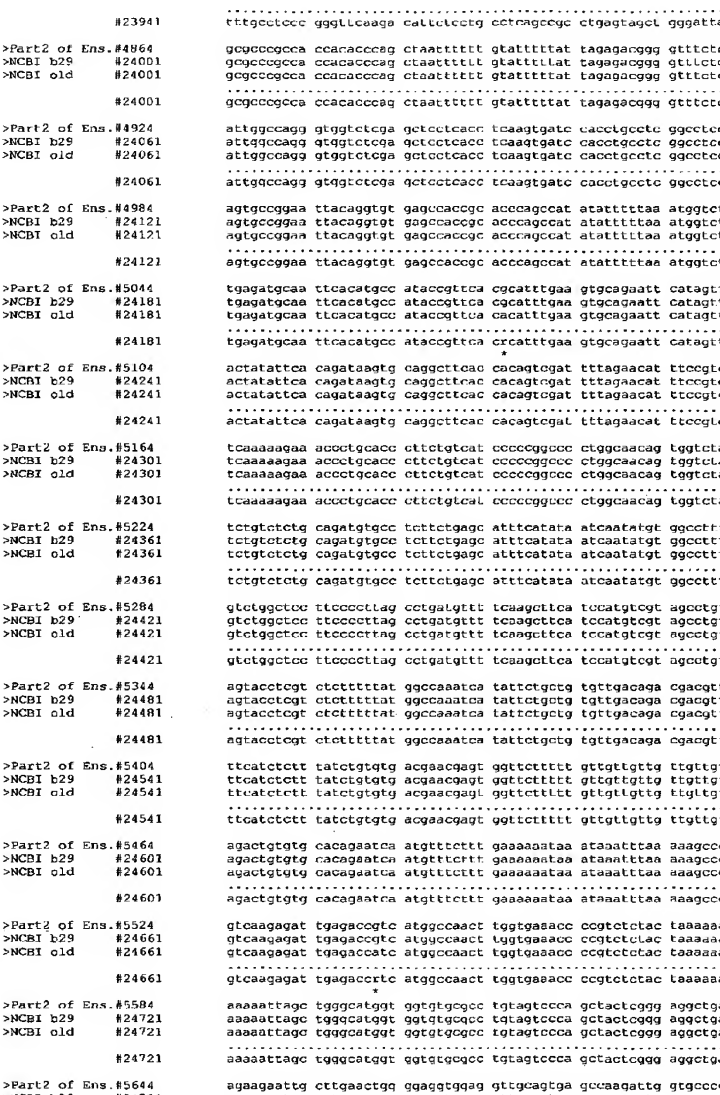


FIGURE 2CC

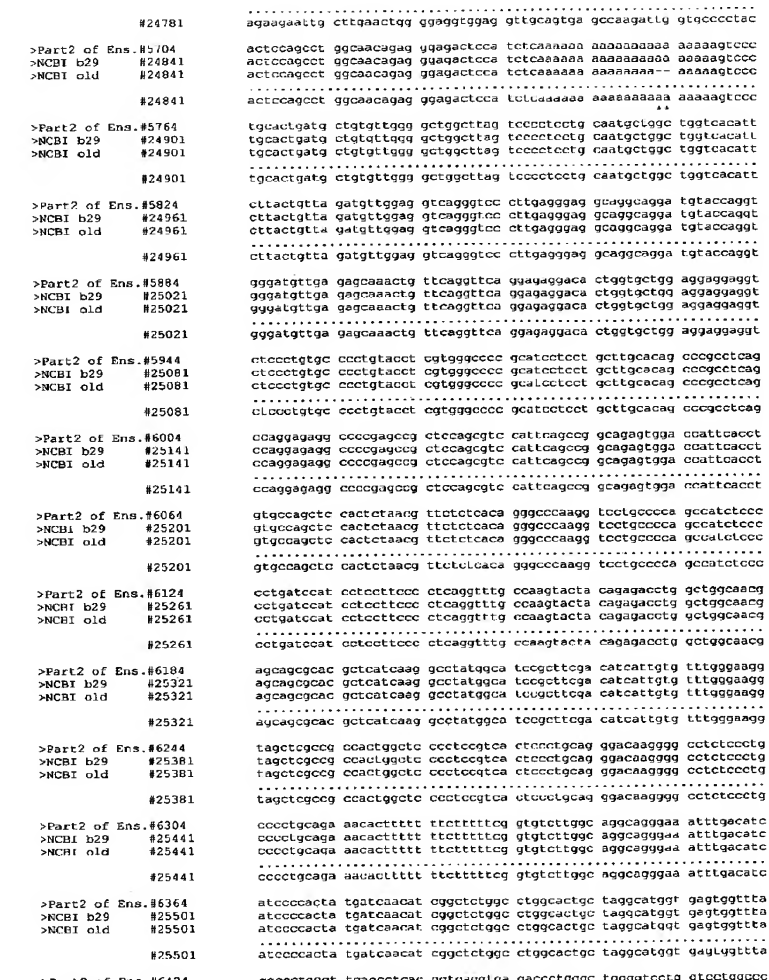


FIGURE 2DD

```
                 ............................................................
        #25621   taggccctag acctcagatg tgtttctaaa cttgaccctc ctaccttctt tccctlggcc

>Part2 of Ens.#6544   cccagccctc ctcccacctt cccttctcca agaccacccc cctcaggtcc cagccttctc
>NCBI b29     #25681   cccagccctc ctcccacctt cccttctcca agaccacccc cctcaggtcc cagccttctc
>NCBI old     #25681   cccagccctc ctcccacctt cccttctcca agaccacccc cctcaggtcc cagccttctc
                 ............................................................
        #25681   cccagccctc ctcccacctt cccttctcca agaccacccc cctcaggtcc cagccttctc

>Part2 of Ens.#6604   ccaagagatg ggagtgcctt tccattccgg taaagattcc aggcttctca ggaaggggca
>NCBI b29     #25741   ccaagagatg ggagtgcctt tccattccgg taaagattcc aggcttctca ggaaggggca
>NCBI old     #25741   ccaagagatg ggagtgcctt tccattccgg taaagattcc aggcttctca ggaaggggca
                 ............................................................
        #25741   ccaagagatg ggagtgcctt tccattccgg taaagattcc aggcttctca ggaaggggca

>Part2 of Ens.#6664   cgcaaagaat aagatgggtt gatgggttgc aagcatcctg gctcactctc accctatgct
>NCBI b29     #25801   cgcaaagaat aagatgggtt gatgggttgc aagcatcctg gctcactctc accctatgct
>NCBI old     #25801   cgcaaagaat aagatgggtt gatgggttgc aagcatcctg gctcactctc accctatgct
                 ............................................................
        #25801   cgcaaagaat aagatgggtt gatgggttgc aagcatcctg gctcactctc accctatgct

>Part2 of Ens.#6724   aaactcaggc gaccgtgctg tgtgacatca tagtcctcta ctgcatgaag aaaagactct
>NCBI b29     #25861   aaactcaggc gaccgtgctg tgtgacatca tagtcctcta ctgcatgaag aaaagactct
>NCBI old     #25861   aaactcaggc gaccgtgctg tgtgacatca tagtcctcta ctgcatgaag aaaagactct
                 ............................................................
        #25861   aaactcaggc gaccgtgctg tgtgacatca tagtcctcta ctgcatgaag aaaagactct

>Part2 of Ens.#6784   actatcggga gaagaaatat aaatatgtgg aagattacga gcaggtaggc ccctcctggc
>NCBI b29     #25921   actatcggga gaagaaatat aaatatgtgg aagattacga gcaggtaggc ccctcctggc
>NCBI old     #25921   actatcggga gaagaaatat aaatatgtgg aagattacga gcaggtaggc ccctcctggc
                 ............................................................
        #25921   actatcggga gaagaaatat aaatatgtgg aagattacga gcaggtaggc ccctcctggc

>Part2 of Ens.#6844   ccccagcagg cacaggcctc tcatctcttg ggtgtgggag ccctgggcgt gggcctgtct
>NCBI b29     #25981   ccccagcagg cacaggcctc tcatctcttg ggtgtgggag ccctgggcgt gggcctgtct
>NCBI old     #25981   ccccagcagg cacaggcctc tcatctcttg ggtgtgggag ccctgggcgt gggcctgtct
                 ............................................................
        #25981   ccccagcagg cacaggcctc tcatctcttg ggtgtgggag ccctgggcgt gggcctgtct

>Part2 of Ens.#6904   ggggaggccc ttctgcagag gctggcacca gtgtggcgtg gtgtccccgt taacccgggc
>NCBI b29     #26041   ggggaggccc ttctgcagag gctggcacca gtgtggcgtg gtgtccccgt taacccgggc
>NCBI old     #26041   ggggaggccc ttctgcagag gctggcacca gtgtggcgtg gtgtccccgt taacccgggc
                 ............................................................
        #26041   ggggaggccc ttctgcagag gctggcacca gtgtggcgtg gtgtccccgt taacccgggc

>Part2 of Ens.#6964   agtcctgcca ctctcagcag ctgctccatc cctaggcccc tgtactagat tgactcttta
>NCBI b29     #26101   agtcctgcca ctctcagcag ctgctccatc cctaggcccc tgtactagat tgactcttta
>NCBI old     #26101   agtcctgcca ctctcagcag ctgctccatc cctaggcccc tgtactagat tgactcttta
                 ............................................................
        #26101   agtcctgcca ctctcagcag ctgctccatc cctaggcccc tgtactagat tgactcttta

>Part2 of Ens.#7024   aacccagcct cgtttcaatg aacgacatct caggttggtg atgataatgc atgctctgag
>NCBI b29     #26161   aacccagcct cgtttcaatg aacgacatct caggttggtg atgataatgc atgctctgag
>NCBI old     #26161   aacccagcct cgtttcaatg agcgacatct caggttggtg atgataatgc atgctctgag
                 ............................................................
        #26161   aacccagcct cgtttcaatg arcgacatct caggttggtg atgataatgc atgctctgag
                                        *
>Part2 of Ens.#7084   aatgcctgtg ggcacacact acttcagtgc accttgcgga acaggaaggg ttgggttcca
>NCBI b29     #26221   aatgcctgtg ggcacacact acttcagtgc accttgcgga acaggaaggg ttgggttcca
>NCBI old     #26221   aatgcctgtg ggcacacact acttcagtgc accttgcgga acaggaaggg ttgggttcca
                 ............................................................
        #26221   aatgcctgtg ggcacacact acttcagtgc accttgcgga acaggaaggg ttgggttcca

>Part2 of Ens.#7144   ggcctgggac caacttgaga acccctggcg gtgaagtccc aggagcgcac ctccctcccg
>NCBI b29     #26281   ggcctgggac caacttgaga acccctggcg gtgaagtccc aggagcgcac ctccctcccg
>NCBI old     #26281   ggcctgggac caacttgaga acccctggcg gtgaagtccc aggagcgcac ctccctcccg
                 ............................................................
        #26281   ggcctgggac caacttgaga acccctggcg gtgaagtccc aggagcgcac ctccctcccg

>Part2 of Ens.#7204   cctgccacaa ggggtcccag gggcaccttg atctgcttgt gtccttcttt gcagggtctt
>NCBI b29     #26341   cctgccacaa ggggtcccag gggcaccttg atctgcttgt gtccttcttt gcagggtctt
>NCBI old     #26341   cctgccacaa ggggtcccag gggcaccttg atctgcttgt gtccttcttt gcagggtctt
                 ............................................................
        #26341   cctgccacaa ggggtcccag gggcaccttg atctgcttgt gtccttcttt gcagggtctt

>Part2 of Ens.#7264   gctagtgagc tggaccagtg aggcctaccc cacacctggg ctctccacag ccccatcaaa
>NCBI b29     #26401   gctagtgagc tggaccagtg aggcctaccc cacacctggg ctctccacag ccccatcaaa
>NCBI old     #26401   gctagtgagc tggaccagtg aggcctaccc cacacctggg ctctccacag ccccatcaaa
                 ............................................................
        #26401   gctagtgagc tggaccagtg aggcctaccc cacacctggg ctctccacag ccccatcaaa

>Part2 of Ens.#7324   gaacagagag gaggaggagg gagaaatggc caccacatca ccccagagaa atttctggaa
>NCBI b29     #26461   gaacagagag gaggaggagg gagaaatggc caccacatca ccccagagaa atttctggaa
>NCBI old     #26461   gaacagagag gaggaggagg gagaaatggc caccacatca ccccagagaa atttctggaa
```

FIGURE 2EE

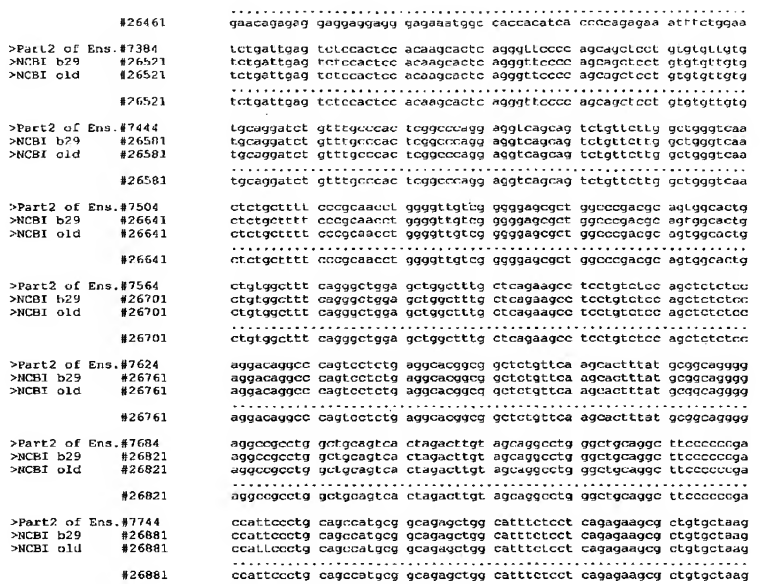


FIGURE 2FF

```
                #27361      cctgacctca aatgatccay ctgccttagc ctctcaaagt gattggatta caggtgtgcq
                                                  *
>NCBI b29       #27421      ccaccacacc tggcctattt attttttta ttacccaggc tggagtgcag tggcgcaatc
>NCBI old       #27421      ccaccacacc tggcctattt attttttta ttacccaggc tggagtgcag tggcgcaatc
                            ..........................................................
                #27421      ccaccacacc tggcctattt attttttta ttacccaggc tggagtgcag tggcgcaatc

>NCBI b29       #27481      actgctcact gcagtctcca cctcctgggc ttaagcgatc ctcctacctc agcctcccat
>NCBI old       #27481      actgctcact gcagtctcca cctcctgggc ttaagcgatc ctcctacctc agcctcccat
                            ..........................................................
                #27481      actgctcact gcagtctcca cctcctgggc ttaagcgatc ctcctacctc agcctcccat

>NCBI b29       #27541      agtgctggga ttacaagcat gagccactgc acccagctgc ctacttttta aagcctaata
>NCBI old       #27541      agtgctggga ttacaagcat gagccactgc acccagctgc ctacttttta aagcctaata
                            ..........................................................
                #27541      agtgctggga ttacaagcat gagccactgc acccagctgc ctacttttta aagcctaata

>NCBI b29       #27601      atttgtcatt agcttgtctt cctgcacagt tcacaggaca ctttttctct agtgacatct
>NCBI old       #27601      atttgtcatt agcttgtctt cctgcacagt tcacaggaca ctttttctct agtgacatct
                            ..........................................................
                #27601      atttgtcatt agcttgtctt cctgcacagt tcacaggaca ctttttctct agtgacatct

>NCBI b29       #27661      tacctgtgcc tccattttga cataaaatta caattctatg ctgtgggtaa cacaattttt
>NCBI old       #27661      tacctgtgcc tccattttga cataaaatta caattctatg ctgtgggtaa cacaattttt
                            ..........................................................
                #27661      tacctgtgcc tccattttga cataaaatta caattctatg ctgtgggtaa cacaattttt

>NCBI b29       #27721      ggtaacatgg aagtggcttc ctctcccttc cctgtttgct tcccaaaagt ctccaaccaa
>NCBI old       #27721      ggtaacatgg aagtggcttc ctctcccttc cctgtttgct tcccaaaagt ctccaaccaa
                            ..........................................................
                #27721      ggtaacatgg aagtggcttc ctctcccttc cctgtttgct tcccaaaagt ctccaaccaa

>NCBI b29       #27781      aagcagcagc acgcatacct agctccgcac aggttcgaaa tgttcatgaa gctatttcca
>NCBI old       #27781      aagcagcagc acgcatacct agctccgcac aggttcgaaa tgttcatgaa gctatttcca
                            ..........................................................
                #27781      aagcagcagc acgcatacct agctccgcac aggttcgaaa tgttcatgaa gctatttcca

>NCBI b29       #27841      gctgcagctt cacaccaact aaatgtttgt gttttttttgt ttttgagaca gagtcttgcc
>NCBI old       #27841      gctgcagctt cacaccaact aaatgtttgt gttttttttgt ttttgagaca gagtcttgcc
                            ..........................................................
                #27841      gctgcagctt cacaccaact aaatgtttgt gttttttttgt ttttgagaca gagtcttgcc

>NCBI b29       #27901      ctattgccca ggctggagtg cagtggcatg atctcggctc actgcaacct ctgcctcctg
>NCBI old       #27901      ctattgccca ggctggagtg cagtggcatg atctcggctc actgcaacct ctgcctcctg
                            ..........................................................
                #27901      ctattgccca ggctggagtg cagtggcatg atctcggctc actgcaacct ctgcctcctg

>NCBI b29       #27961      ggttcaagca attctcctgc ctcagcctcc caagtagctg ggattacagg catctgccac
>NCBI old       #27961      ggttcaagca attctcctgc ctcagcctcc caagtagctg ggattacagg catctgccac
                            ..........................................................
                #27961      ggttcaagca attctcctgc ctcagcctcc caagtagctg ggattacagg catctgccac

>NCBI b29       #28021      cacacccagc taatttttgt atttttagta gagacagagg ttcaccatgt tggccaggct
>NCBI old       #28021      cacacccagc taatttttgt atttttagta gagacagagg ttcaccatgt tggccaggct
                            ..........................................................
                #28021      cacacccagc taatttttgt atttttagta gagacagagg ttcaccatgt tggccaggct

>NCBI b29       #28081      ggtctggaac tcctgacctc atgatccacc gcctcagcct cccaaagtgc tgggattaca
>NCBI old       #28081      ggtctggaac tcctgacctc atgatccacc gcctcagcct cccaaagtgc tgggattaca
                            ..........................................................
                #28081      ggtctggaac tcctgacctc atgatccacc gcctcagcct cccaaagtgc tgggattaca

>NCBI b29       #28141      ggcgtgagcc actgcaccca gccacacacc agctgaattc tgcatcaggt attcagaatc
>NCBI old       #28141      ggcgtgagcc actgcaccca gccacacacc agctgaattc tgcatcaggt attcagaatc
                            ..........................................................
                #28141      ggcgtgagcc actgcaccca gccacacacc agctgaattc tgcatcaggt attcagaatc

>NCBI b29       #28201      cccctagagg aagcagtttg cccttcgcct ctgccatgaa ttcacttcca agtactgaga
>NCBI old       #28201      cccctagagg aagcagtttg cccttcgcct ctgccatgaa ttcacttcca agtactgaga
                            ..........................................................
                #28201      cccctagagg aagcagtttg cccttcgcct ctgccatgaa ttcacttcca agtactgaga

>NCBI b29       #28261      ggaaccatta ttttacaaat actctgcttc tgctggataa atggccaacc tgggcttttc
>NCBI old       #28261      ggaaccatta ttttacaaat actctgcttc tgctggataa atggccaacc tgggcttttc
                            ..........................................................
                #28261      ggaaccatta ttttacaaat actctgcttc tgctggataa atggccaacc tgggcttttc

>NCBI b29       #28321      ttgcttttgg aaaaaaaaaa aaaaaaagtt taccaactcc ctttattgaa gctctggttc
>NCBI old       #28321      ttgcttttgg ---aaaaaaa aaaaaaagtt taccaactcc ctttattgaa gctctggttc
                            ..........................................................
                #28321      ttgcttttgg aaaaaaaaaa aaaaaaagtt taccaactcc ctttattgaa gctctggttc
                                       ***

>NCBI b29       #28381      atgctgcctt ttaaaaaaaa atgcttgttg cctgtaatcc caggactttg ggaggccaag
>NCBI old       #28381      atgctgcctt ttaaaaaaaa atgcttgttg cctgtaatcc caggactttg ggaggccaag
```

FIGURE 2GG

```
                                ....................................................................
                #28381          atgctgcctt ttaaaaaaaa atgcttgttg cctgtaatcc caggactttg ggaggccaag

>NCBI b29       #28441          gtggatcatt taagcgtagg cattcaagac cagcccagtc aatatggtga gaccctatct
>NCBI old       #28441          gtggatcact taagcgtagg cattcaagac cagcccagtc aacatggtga gaccctatct
                                ....................................................................
                #28441          gtggatcayt taagcgtagg cattcaagac cagcccagtc aayatggtga gaccctatct
                                 *                                                  *
>NCBI b29       #28501          ctacaaaaaa ttaaattagc cgggtgtggc ggtatatatc tgtagtccca gctacttggg
>NCBI old       #28501          ctacaaaaaa ttaaattagc cgggtgtggc ggtatatatc tgtagtccca gctacttggg
                                ....................................................................
                #28501          ctacaaaaaa ttaaattagc cgggtgtggc ggtatatatc tgtagtccca gctacttggg

>NCBI b29       #28561          tggctgaggc cagaggatca cttgagccca ggaggttgag gctgcagtga gctatg
>NCBI old       #28561          tggctgaggc cggaggatca cttgagccca ggaggttgag gctgcagtga gctatg
                                ....................................................................
                #28561          tggctgaggc crgaggatca cttgagccca ggaggttgag gctgcagtga gctatg
                                          *
```

FIGURE 2HH

MAGCCAALAAFLFEYDTPRIVLIRSRKVGLMNRAVQLLILAYVIGWVFVWEKGYQETDSVVS
SVTTKVKGVAVTNTSKLGFRIWDVADYVIPAQEENSLFVMTNVILTMNQTQGLCPEIPDATT
VCKSDASCTAGSAGTHSNGVSTGRCVAFNGSVKTCEVAAWCPVEDDTHVPQPAFLKAAENFT
LLVKNNIWYPKFNFSKRNILPNITTTYLKSCIYDAKTDPFCPIFRLGKIVENAGHSFQDMAV
EGGIMGIQVNWDCNLDRAASLCLPRYSFRRLDTRDVEHNVSPGYNFRFAKYYRDLAGNEQRT
LIKAYGIRFDIIVFGKAGKFDIIPTMINIGSGLALLGMATVLCDIIVLYCMKKRLYYREKKY
KYVEDYEQGLASELDQ

FIGURE 3A

MTNVILTMNQTQGLCPEIPDATTVCKSDASCTAGSAGTHSNGVSTGRCVAFNGSVKTCEVAA
WCPVEDDTHVPQPAFLKAAENFTLLVKNNIWYPKFNFSKRNILPNITTTYLKSCIYDAKTDP
FCPIFRLGKIVENAGHSFQDMAVEGGIMGIQVNWDCNLDRAASLCLPRYSFRRLDTRDVEHN
VSPGYNFRFAKYYRDLAGNEQRTLIKAYGIRFDIIVFGKAGKFDIIPTMINIGSGLALLGMA
TVLCDIIVLYCMKKRLYYREKKYKYVEDYEQGLASELDQ

FIGURE 3B

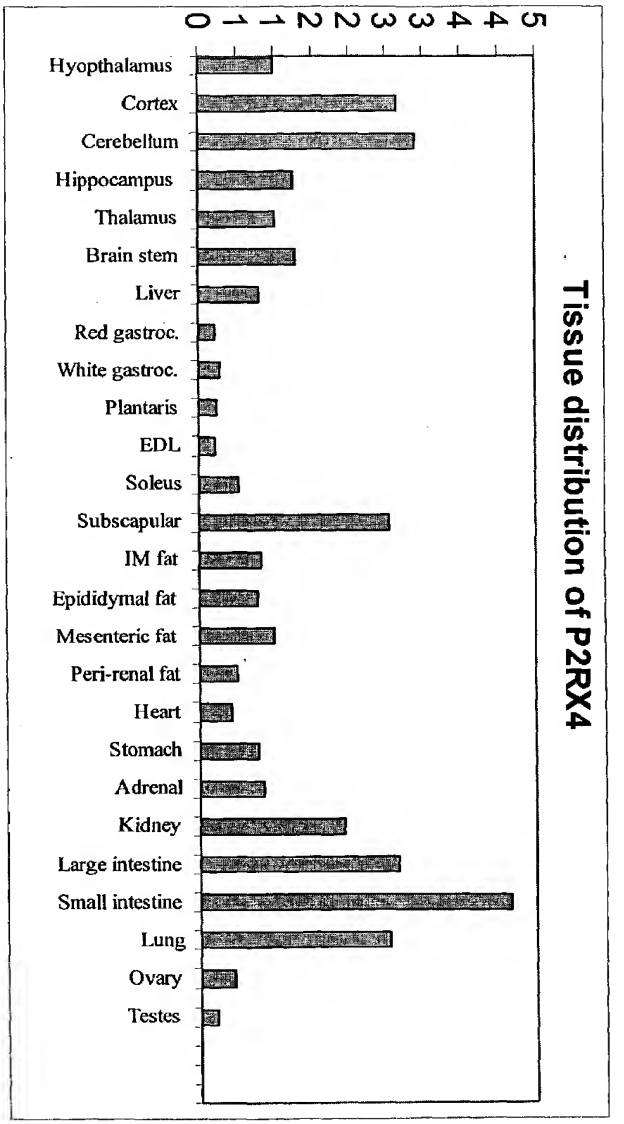

Tissue distribution of P2RX4
0
1
1
2
2
3
3
4
4
5
Hyopthalamus
Cortex
Cerebellum
Hippocampus
Thalamus
Brain stem
Liver
Red gastroc.
White gastroc.
Plantaris
EDL
Soleus
Subscapular
IM fat
Epididymal fat
Mesenteric fat
Peri-renal fat
Heart
Stomach
Adrenal
Kidney
Large intestine
Small intestine
Lung
Ovary
Testes

FIGURE 4

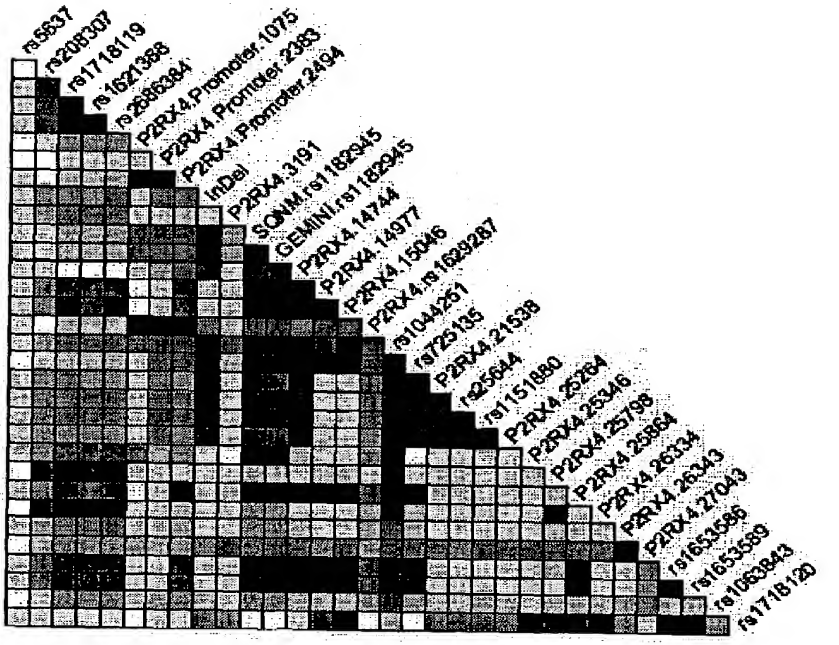
gcctagttgtactcacagcattgtggagattggcttccaaggacatttggggttctgtta

ccagaaaagctgggagcagatgctgggaagaaaacaacggatgtttgctacagtgtattt

taaactaaactagacaggttcatcctggcgcgg

FIGURE 5

FIGURE 6


rs5637
rs208307
rs1718119
rs1621388
rs2885384
P2RX4.Promoter.1075
P2RX4.Promoter.2383
P2RX4.Promoter.2494
InDel
P2RX4.3191
SONM.rs1182945
GEMINI.rs1182945
P2RX4.14744
P2RX4.14977
P2RX4.15046
P2RX4.rs1629267
rs1044261
rs725135
P2RX4.21538
rs25644
rs1151680
P2RX4.25264
P2RX4.25346
P2RX4.25798
P2RX4.25864
P2RX4.26334
P2RX4.26343
P2RX4.27049
P2RX4.27686
rs1653586
rs1653589
rs1062843
rs1718120

DIAGNOSING PREDISPOSITION TO FAT DEPOSITION AND THERAPEUTIC METHODS FOR REDUCING FAT DEPOSITION AND TREATMENT OF ASSOCIATED CONDITIONS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims benefit of U.S. Serial No. 60/386,012 filed Jun. 4, 2002. The contents of that application is incorporated herein by reference.

FIELD OF THE INVENTION

[0002] The invention relates to genetic alterations in nucleic acids that are associated with fat deposition and diabetes.

BACKGROUND

[0003] Individuals who are obsese have excess body fat compared to set standards. Obesity can be determined by several methods including body mass index (BMI) measurements, weight-for height charts, and body fat measurements determined by skinfold thickness and bioelectrical impedance. Obesity affects 58 million people across the United States, which represents approximately one-quarter to one-third of the adult population, and its prevalence is increasing to epidemic proportions in the United States and in other industrialized nations.

[0004] Recognized since 1985 as a chronic disease, obesity-related medical conditions contribute to approximately 300,000 deaths each year, second only to smoking as a cause of preventable death. (JAMA, 276: 1907-1915 (1996)). Obesity has been established as a major risk factor for type II diabetes melitus, hypertension, cardiovascular disease and some cancers in both men and women (JAMA, 282: 1523-1529 (1999)). Other comorbid conditions include sleep apnea, osteoarthritis, infertility, idiopathic intracranial hypertension, lower extremity venous stasis disease, gastro-esophageal reflux and urinary stress incontinence.

[0005] The total cost attributable to obesity amounted to $99.2 billion in 1995. Approximately $51.65 billion of those dollars were direct medical costs. The cost of obesity to U.S. business in 1994 was estimated to total $12.7 billion, and health-related economic costs of obesity to businesses in the United States is substantial, representing approximately 5% of total medical care costs. (American Journal of Health Promotion, 13 (2): 120-127 (1998)). It was found that as BMI increases, so do the number of sick days, medical claims and health care costs and that the mean annual health care costs for the BMI "at risk" population was $2,274 versus $1,499 for the "not at risk" group.

[0006] An accumulation of adipose tissue on the trunk and around the waist, known as central fat, also confers an increased risk of type II diabetes and cardiovascular disease (Lundgren et al., *Int. J. Obes.,* 13(4): 413-23 (1989); Ohlson et al., *Diabetes,* 34(10): 1055-8 (1985)). In addition, central obesity has been implicated in a condition known as the metabolic syndrome (or syndrome X), which is associated with increased risk of cardiovascular disease, vascular dementia, and diabetes. The metabolic syndrom is a descriptive term for the coexistence of all of the following or differing combinations of central fat, hypertension, glucose intolerance, dyslipidemia (elevated triglycerides and low HDL cholesterol), and impaired insulin stimulated glucose uptake ("insulin resistance"). Prevalence of central fat and its relationship to general obesity differs between ethnic groups and gender (McKeigue et al., *Diabetologia,* 35(8): 785-91 (1992); McKeigue et al., *Lancet,* 337(8738): 382-6 (1991)). A majority of male subjects having high central fat are also obese in terms of BMI, and obese subjects often have a central distribution of fat, which suggests an overlap between these two conditions. While this relationship is not as strongly correlated in women, central fat increases after menopause.

[0007] Current anti-obesity therapeutics (e.g., Phentermine, Sibutramine, and Orlistat) are largely ineffective and there is an urgent need to define the etiology of this disease and initiate rational, mechanism-based drug development. Mouse QTL and human studies have postulated that the 12q22 to q23 region, and specifically the insulin-like growth factor 1 (IGF-1) gene in that region, play a role in body weight regulation and visceral fat deposition (Collins, A. C. et al., *Mamm. Genome,* 4: 454-458 (1993); Sun, G. et al., *Int. J. Obes.,* 23: 929-935 (1999); Keightley, P. D. et al., *Genetics,* 142: 227-235 (1996). Also, other studies have linked obesity with certain portions of the human genome (Perusse, L. et al., *Obesity Research,* 9: 135-169 (2001); Chagnon, Y. C. et al., *Obesity Research,* 8: 89-117 (2000)). Specifically, the CD36L gene on chromosome 12 was implicated in plasma lipid levels and with BMI (Acton, S. et al., *Arterioscler. Thromb. Vasc. Biol.,* 19: 1734-1743 (1999)), the 12q24 chromosomal region was postulated as playing a role in obesity in a Quebec Family Study (Perusse, L. et al., *Diabetes,* 50: 614-621 (2001)), and it was reported that certain polymorphic loci on chromosome four are associated with obesity (Stone et al., *American J Human Genetics,* "A major predisposition locus for severe obesity at 4p15-p14," June 2002).

SUMMARY

[0008] It has been discovered that polymorphic variations in a gene encoding a purinergic receptor known as P2X4, which is located on chromosome twelve, have been associated with central fat deposition. Thus, featured herein is a method for diagnosing predisposition to fat deposition or leanness in a subject, which comprises detecting the presence or absence of one or more polymorphic variations in a P2X4 nucleotide sequence in a nucleic acid sample from a subject, where the P2X4 nucleotide sequence is set forth as SEQ ID NO: 1 or a substantially identical nucleotide sequence thereof. In certain embodiments, polymorphic variations at positions 11030, 15847, or 17338 in SEQ ID NO: 1 are detected.

[0009] Also featured herein are nucleic acids that encode a P2X4 polypeptide and include one or more polymorphic variations at positions 11030, 15847, or 17338 of SEQ ID NO: 1, and oligonucleotides which hybridize to those nucleic acids. Also featured are methods for identifying candidate therapeutic molecules that reduce fat deposition and treat related disorders, as well as methods of reducing fat deposition and treating related disorders in a subject by administering such a therapeutic molecule.

[0010] In addition, it was discovered that a polymorphic variation at position 15847 was associated with type II

diabetes (non-insulin dependent diabetes mellitus, or NIDDM) in subjects. Therefore, featured herein is a method for diagnosing predisposition to NIDDM in a subject, which comprises obtaining a nucleic acid sample from the subject and detecting the presence or absence of a polymorphic variation in a P2X4 nucleotide sequence associated with NIDDM. Also featured are methods for identifying candidate therapeutic molecules for treating NIDDM in a subject and methods for treating NIDDM in a subject by administering such a therapeutic molecule.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] FIGS. 1A-1L depicts the P2X4 nucleotide sequence reported as SEQ ID NO:1. The following nucleotide representations are used throughout: "A" or "a" is adenosine, adenine, or adenylic acid; "C" or "c" is cytidine, cytosine, or cytidylic acid; "G" or "g" is guanosine, guanine, or guaylic acid; "T" or "t" is thymidine, thymine, or thymidylic acid; and "I" or "i" is inosine, hypoxanthine, or inosinic acid. Exons are indicated in italicized lower case type, introns are depicted in normal text lower case type, and polymorphic sites are depicted in bold upper case type. The sequence designates polymorphic sites at positions 2878, 11030, 15847, and 17338, where a GGGCCCC nucleotide sequence insert was identified within positions 2878 to 2891 in a minority of samples (not shown in the figure), and where single nucleotide polymorphisms (SNPs) are indicated at positions 11030, 15847, and 17338. SNPs are designated by the following convention: "R" represents A or G, "M" represents A or C; "W" represents A or T; "Y" represents C or T; "S" represents C or G; "K" represents G or T; "V" represents A, C or G; "H" represents A, C, or T; "D" represents A, G, or T; "B" represents C, G, or T; and "N" represents A, G, C, or T. There are three transcription products and two translation products that can be expressed from the nucleotide sequence as it is a dicistronic sequence. The first transcription product is encoded by a nucleic acid that starts at position 3027 and ends at position 36419, where exon 1 begins at position 2719 and ends at position 3160, exon 2 beings at position 9997 and ends at position 10144, exon 3 begins at position 14776 and ends at position 14847, exon 4 begins at position 14957 and ends at position 15029, exon 5 begins at position 15810 and ends at position 15906, exon 6 begins at position 21398 and ends at position 21478, exon 7 begins at position 21590 and ends at position 21731, exon 8 begins at position 21828 and ends at position 21964, exon 9 begins at position 25286 and ends at position 25379, exon 10 begins at position 25483 and ends at position 25548, exon 11 begins at position 25869 and ends at position 25964, and exon 12 begins at position 26395 and ends at position 26616. The second transcription product is encoded by a nucleic acid that also starts at position 3027 and ends at position 26419, where exon 1 begins at position 3001 and ends at position 3160, exon 2 beings at position 9997 and ends at position 10144, exon 3 begins at position 14776 and ends at position 14847, exon 4 begins at position 14957 and ends at position 15029, exon 5 begins at position 15810 and ends at position 15906, exon 6 begins at position 21398 and ends at position 21478, exon 7 begins at position 21590 and ends at position 21731, exon 8 begins at position 21828 and ends at position 21964, exon 9 begins at position 25286 and ends at position 25379, exon 10 begins at position 25483 and ends at position 25548, exon 11 begins at position 25869 and ends at position 25964, and exon 12 begins at position 26395 and ends at position 26616. The third transcription product is encoded by a nucleic acid which begins at position 14797 and ends at position 16419, where exon 1 begins at position 3129 and ends at position 3160, exon 2 beings at position 9997 and ends at position 10144, exon 3 begins at position 14776 and ends at position 14847, exon 4 begins at position 14957 and ends at position 15029, exon 5 begins at position 15810 and ends at position 15906, exon 6 begins at position 21398 and ends at position 21478, exon 7 begins at position 21590 and ends at position 21731, exon 8 begins at position 21828 and ends at position 21964, exon 9 begins at position 25286 and ends at position 25379, exon 10 begins at position 25483 and ends at position 25548, exon 11 begins at position 25869 and ends at position 25964, and exon 12 begins at position 26395 and ends at position 26974.

[0012] FIGS. 2A-2HH depicts an alignment between nucleotide sequences reported in NCBI and Ensembl databases used to determine the nucleotide sequence of SEQ ID NO: 1.

[0013] **FIG. 3A** shows a polypeptide sequence encoded by the first translation start site at position 3027 of SEQ ID NO:1 and is a translation product of the first or second transcription product. The polypeptide in **FIG. 3A** is 87% identical to *R. norvegicus* P2X4, which is 387 amino acids in length (sp:P51577) and 80% identical to *M. musculus* P2X4, which is also 387 amino acids in length (ref:NP_035156.1, DNA segment, Chr 5, ERATO Doi 444). **FIG. 3B** shows a polypeptide sequence encoded by the third transcription product, which begins after the first ribosome entry site.

[0014] **FIG. 4** depicts a tissue expression profile for P2X4.

[0015] **FIG. 5** shows a nucleotide sequence insert (SEQ ID NO:) that is optionally included in SEQ ID NO:1 between positions 19137 and 19138.

[0016] **FIG. 6** shows a plot of pairwise linkage disequilibrium estimates D' and pairwise linkage disequilibrium estimates for polymorphisms identified in and around the P2X4 gene. Dark colored squares represent high linkage disequilibrium (LD) between SNPs, and no color or light colored squares represent low LD between SNPs.

DETAILED DESCRIPTION

[0017] It has been discovered that polymorphic variants in or near a gene on chromosome 12 encoding a purinergic receptor are associated with fat deposition in the abdomen and trunk region of subjects. Individuals having increased fat deposition in this area are at risk of developing metabolic conditions (e.g., diabetes and obesity) and cardiovascular conditions (e.g., hypertension). Thus, methods for detecting genetic determinants for fat deposition can lead to early diagnosis of a predisposition to these conditions (e.g., hyperinsulinaemia, hypertension, glucose intolerance (that is, IGT or diabetes), dyslipidemia, hypercoagulability and microalbuminuria) and early prescription of preventative measures. Also, associating P2X4 with fat deposition has provided a new target for screening molecules useful for treatments that reduce fat deposition. P2X4 is also a target for screening molecules useful for treating disorders associated with fat deposition, which include metabolic disorders (e.g., diabetes and obesity) and cardiovascular disorders (e.g., hypertension).

[0018] Purinergic receptors are polypeptides located in cell membranes capable of binding adenosine diphosphate (ADP) and adenosine triphosphate (ATP). These receptors regulate biological processes including neurotransmission, muscle contraction, and endocrine secretion. There are two families of purinogenic receptors: the P2Y family of G-protein coupled receptors the P2X family of ligand-gated cation channels. P2X4 is one of seven members in the P2X family of purinergic receptors, where activation of this receptor leads to opening of non-selective cation channels permeable to Na^+, K^+ or Ca^{2+} (Xiong et al., *Br. J. Pharmacol,* 130(6): 1394-1398 (2000)). P2X4 cDNA has been isolated from human brain and mapped to chromosomal locus 12q24.32. This cDNA encodes a polypeptide having 87% sequence identity to rat P2X4. Splice variants of human P2X4 (Dhulipala et al., *Gene,* 207: 259-266 (1998)) and mouse P2X4 (Townsend-Nicholoson et al., *Brain Res. Mol. Brain Res.,* 64: 246-254 (1999)) have been identified.

[0019] P2X4 is expressed in neuronal and non-neuronal tissues, and expression was shown in the hippocampus of the brain and in pancreatic islets in the rat (Wang et al., *Biochem. Biophys. Res. Commun.,* 220: 196-202 (1996)). Also, P2X4 is localized in adrenal cortical cells of the zona reticularis in adult rats, and it has been hypothesised that circadian variation and corticosteroid secretion by the adrenal cortex may effect expression levels of purinergic receptors (Afework and Burnstock, *Int. J. Devl. Neuroscience,* 18: 515-520 (2000)). P2X4 has been distinguished from other P2X members on the basis of its relative insensitivity to many P2X receptor antagonists (Buell et al., *EMBO J.,* 15: 55-62 (1995); Collo et al., *Journal of Neuroscience,* 16: 2495-2507 (1996)). It was reported that ethanol inhibits P2X4 function in Xenopus oocytes, suggesting that P2X4 may play a role in the action of ethanol on the central nervous system (Kiefer et al., *Alcohol Clin. Exp. Res.,* 25(5): 787-789 (2001)).

[0020] Central Fat Deposition and Associated Conditions

[0021] Many individuals considered as having increased central fat deposition are also considered obese according to BMI, weight-for-height charts, or body fat measurements. Obesity is generally understood as a condition where fat content in an individual is above a predetermined level. For example, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health (NM, see http address www.nih.gov/health/nutrit/pubs/statobes.htm) define individuals having a body mass index (BMI) of 25 to 29.9 kg/m^2 as being overweight and individuals having a BMI of 30 kg/m^2 or greater as being obese.

[0022] Increased central fat levels also have been linked to the metabolic syndrome, which includes the coexistence or one or more life threatening medical conditions such as metabolic conditions (e.g., diabetes and obesity) and cardiovascular conditions (e.g., myocardial infarction and hypertension). For example, cardiovascular mortality was assessed in 3,606 subjects from the Botnia study (a large-scale study of type 2 diabetes begun in Finland in 1990) with a median follow-up of 6.9 years. In women and men, respectively, the metabolic syndrome was recorded in 10 and 15% of subjects with normal glucose tolerance, 42 and 64% of those with IFG/IGT, and 78 and 84% of those with type 2 diabetes. The risk for coronary heart disease and stroke was increased threefold in subjects with the syndrome, and cardiovascular mortality was markedly increased (12.0% in subjects with the syndrome versus 2.2% in those without; $P<0.001$) (Zimmet, et al. (2001) *Nature* 414: 782-787). Thus, determining a predisposition to fat deposition, and specifically central fat deposition, is useful for determining whether a person should be considered for being placed on a preventative regimen for reducing fat, thereby reducing the probability that the person develops one or more conditions linked to fat deposition.

[0023] The term "fat deposition" as used herein refers to fat content in an individual as well as processes in which fat is deposited in certain locations of an individual. The term "central fat deposition" as used herein refers to fat around the trunk and waist of an individual that is above a predetermined level or average in a population. The central region may be defined as the region extending from the superior surface of the second lumbar vertebra extending inferiorly to the inferior surface of the fourth lumbar vertebra and laterally to the inner aspect of the ribcage. Fat deposition can be measured as a quantity at one time point or a quantity over a series of time points, for example, and fat deposition can be quantified or estimated using a number of procedures described hereafter. Fat is composed of adipose cells deposited below the skin (i.e., subcutaneous adipose cells) and/or deeper within an individual's body (i.e., visceral adipose cells). Adipose cells often are connective tissue cells specialized for synthesis and storage of fat. Such cells often contain globules of triglycerides where the nucleus is generally displaced to one side of the globule and the cytoplasm is visualized as a thin line around the fat droplet. Provided herein are methods for detecting predisposition to overall adipose cell deposition in a. subject (i.e., includes subcutaneous adipose cells and visceral adipose cells), as well as methods for distinguishing between a predisposition to subcutaneous adipose cell deposition and a predisposition to visceral adipose cell deposition.

[0024] Fat deposition is quantified in a number of manners (see, e.g., Wajchenberg, *Endocrine Rev.* 21(6): 697-738 (2000)). For example, caliper measurements of skinfold thickness in defined areas of the body have been utilized to differ between different kinds of regional fat (Nordhamn, et al., *Int. J. Obes. Relat. Metab. Disord.* 24(5): 652-7 (2000)). Waist and hip measurements using tape measures are commonly utilized indices of central fat (Lundgren et al., *Int. J. Obes.,* 13(4): 413-23 (1989); Ohlson et al., *Diabetes* 34(10): 1055-8 (1985)), and sagittal abdominal diameter is measured by some researchers for quantifying central fat. Also, computed tomography and X-ray based methods have been utilized to quantify central fat content. Dual x-ray absorbtiometry (DEXA) is relatively fast and inexpensive and yields reliable estimations of body composition (fat mass/lean mass/bone) with reproducibility. DEXA measurements and waist and hip measurements were utilized for quantifying central fat in Example 1. Magnetic resonance imaging (MRI) and computed tomography procedures can be used to distinguish between visceral fat deposition and subcutaneous fat deposition (see e.g. Wajchenberg, supra).

[0025] Thus, fat deposition can be expressed in units used for quantifying fat content. Fat deposition can be expressed in terms of total fat content in an individual or region of an individual (grams or percentage of total weight of an individual), visceral fat content in an individual or region of an

individual (grams, percentage of total weight of an individual, or percentage of total fat in an individual), and subcutatnous fat content in an individual or region of an individual (grams, percentage of total weight of an individual, or percentage of total fat in an individual). Each of these expressions of fat deposition can be measured or quantified at a single point in time or over two or more points in time.

[0026] Fat deposition also can be expressed in terms of "increased fat deposition" (also referred to as "higher fat deposition" and "at increased risk for fat deposition"), which is relative to average fat deposition in a population. In a distribution of fat deposition across a population (expressed in any of the units of measure described herein), individuals having increased fat deposition are sometimes represented in the upper 40% or upper 30% of the population, often in the upper 25%, upper 20%, upper 15%, and upper 10% of the population, and sometimes in the upper 5% of the population. Also, individuals having increased fat deposition can be characterized as having waist/hip ratios of 1.01 or more for males and 0.91 or more for females. In addition, men or women having a BMI between 25 and 30 or between about 1335 and about 2050 grams of central fat are typically considered overweight, and individuals having a BMI over 30 or over about 2050 grams of central fat are normally considered obese (e.g., grams of central fat can be determined by DEXA, as descibed above). Also, "leanness" or "decreased fat deposition" (also referred to as "lower fat deposition" and "at decreased risk for fat deposition") are terms that refer to fat deposition and are also relative to average fat deposition in a population. In a distribution of fat deposition across a population, lean individuals are sometimes represented in the lower 40% or lower 30% of the population, often in the lower 25%, lower 20%, lower 15%, and lower 10% of the population, and sometimes in the lower 5% of the population. Also, lean individuals can be characterized as having waist/hip ratios of 1.00 or less for males and 0.90 or less for females. In addition, men or women having a BMI of 24 or less or less than about 1334 grams of central fat are normally considered lean.

[0027] The term "metabolic condition" as used herein refers to a disease, disorder, or state involving increased or decreased metabolites relative to a population average. Examples of metabolic disorders include but are not limited to diabetes, obesity, anorexia nervosa, cachexia, and lipid disorders. The new paradigm relating to central obesity and type 2 diabetes also influences contemporary therapy for the disease. Evidence now exists for a far more aggressive approach to treating not just the hyperglycaemia, but also other cardiovascular disease risk factors such as hypertension, dyslipidaemia and central obesity in type 2 diabetic patients, with the hope of significantly reducing cardiovascular morbidity and mortality.

[0028] The term "NIDDM" as used herein refers to non-insulin-dependent diabetes mellitus or Type 2 diabetes (the two terms are used interchangeably throughout this document). NIDDM refers to an insulin-related disorder in which there is a relative disparity between endogenous insulin production and insulin requirements, leading to elevated hepatic glucose production, elevated blood glucose levels, inappropriate insulin secretion, and peripheral insulin resistance.

[0029] The term "cardiovascular condition" as used herein refers to a disease, disorder, or state involving the cardiovascular system, e.g., the heart, the blood vessels, and/or the blood. A cardiovascular disorder can be caused by an imbalance in arterial pressure, a malfunction of the heart, or an occlusion of a blood vessel (e.g., by a thrombus). Other examples of cardiovascular disorders include but are not limited to hypertension, atherosclerosis, coronary artery spasm, coronary artery disease, arrhythmias, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts (late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts), early cyanosis (e.g., tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection), obstructive congenital anomalies (e.g., coarctation of aorta, pulmonary stenosis and atresia, and aortic stenosis and atresia), disorders involving cardiac transplantation, and congestive heart failure.

[0030] Polymorphic Variants Associated with Fat Deposition

[0031] A genetic analysis provided herein linked fat deposition with polymorphic variants of a nucleotide sequence located on chromosome twelve that encodes a purinergic receptor designated P2X4. An additional genetic analysis provided herein associated NIDDM with genetic variability of the P2X4 nucleotide sequence. As used herein, the term "polymorphic site" refers to a region in a nucleic acid at which two or more alternative nucleotide sequences are observed in a significant number of nucleic acid samples from a population of individuals. A polymorphic site may be a nucleotide sequence of two or more nucleotides, an inserted nucleotide or nucleotide sequence, a deleted nucleotide or nucleotide sequence, or a microsatellite, for example. A polymorphic site that is two or more nucleotides in length may be 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more, 20 or more, 30 or more, 50 or more, 75 or more, 100 or more, 500 or more, or about 1000 nucleotides in length,

where all or some of the nucleotide sequences differ within the region. A polymorphic site often is one nucleotide in length, which is referred to herein as a "single nucleotide polymorphism" or a "SNP."

[0032] Where there are two, three, or four alternative nucleotide sequences at a polymorphic site, each nucleotide sequence is referred to as a "polymorphic variant." Where two polymorphic variants exist, for example, the polymorphic variant represented in a minority of samples from a population is sometimes referred to as a "minor allele" and the polymorphic variant that is more prevalently represented is sometimes referred to as a "major allele." Many organisms possess a copy of each chromosome (e.g., humans), and those individuals who possess two major alleles or two minor alleles often are referred to as being "homozygous" with respect to the polymorphism and those individuals who possess one major allele and one minor allele are normally referred to as being "heterozygous" with respect to the polymorphism. Individuals who are homozygous with respect to one allele are sometimes predisposed to a different phenotype as compared to individuals who are heterozygous or homozygous with respect to another allele. As used herein, the term "phenotype" refers to a trait which can be compared between indviduals, such as presence or absence of a condition, a visually observable difference in appearance between individuals, metabolic variations, physiological variations, variations in the function of biological molecules, and the like. Examples of phenotypes are fat deposition, obesity, and diabetes.

[0033] Researchers sometimes report a polymorphic variant in a database without determining whether the variant is represented in a significant fraction of a population. Because a subset of these reported polymorphmic variants are not represented in a statistically significant portion of the population, some of them are sequencing errors and/or not biologically relevant. Thus, it often is not known whether a reported polymorphic variant is statistically significant or biologically relevant until the presence of the variant is detected in a population of individuals and the frequency of the variant is determined. Methods for detecting a polymorphic variant in a population is described herein, specifically in Example 2. A polymorphic variant is statistically significant and often biologically relevant if it is represented in 5% or more of a population, sometimes 10% or more, 15% or more, or 20% or more of a population, and often 25%, 30% or more, 35% or more, 40% or more, 45% or more, or 50% or more of a population.

[0034] A polymorphic variant may be detected on either or both strands of a double-stranded nucleic acid. Also, a polymorphic variant may be located within an intron or exon of a gene or within a portion of a regulatory region such as a promoter, a 5' untranslated region (UTR), a 3 UTR, and in DNA (e.g., genomic DNA (gDNA) and complementary DNA (cDNA)), RNA (e.g., mRNA, tRNA, and rRNA), or a polypeptide. Polymorphic variations may or may not result in detectable differences in gene expression, polypeptide structure, or polypeptide function.

[0035] In the genetic analysis that associated polymorphic variations in P2X4 with fat deposition and NIDDM, samples from individuals in a population of twin pairs were genotyped, although other populations could be subjected to analysis. The term "genotyped" as used herein refers to a process for determining a genotype of one or more individuals, where a "genotype" is a representation of polymorphic variants in a population. Fat deposition was quantified in the central region of individuals in the study group, and SNPs were identified at positions 11030, 15847, and 17338 in the P2X4 nucleotide sequence represented by SEQ ID NO: 1. It was determined that 33% of the individuals tested in the genetic analysis had a thymine at position 11030 and 64% of the indivduals had an adenine at this position. At position 15847, 55% of the individuals had a thymine and 45% of the individuals had a cytosine, and at position 17338, 76% of the individuals had a cytosine and 24% of the individuals had a thymine. It was determined that an adenine at position 11030, a thymine at position 15847, or a cytosine at position 17338 of SEQ ID NO:1 were individually associated with central fat deposition, and the presence of a thymine at position 11030, a cytosine at position 15847, or a thymine at position 17338 of SEQ ID NO:1 were individually associated with leanness.

[0036] Furthermore, a genotype or polymorphic variant may be expressed in terms of a "haplotype," which as used herein refers to two or more polymorphic variants occurring within genomic DNA in a group of individuals within a population. For example, two SNPs may exist within a gene where each SNP position includes a cytosine variation and an adenine variation. Certain individuals in a population may carry one allele (heterozygous) or two alleles (homozygous) having the gene with a cytosine at each SNP position. As the two cytosines corresponding to each SNP in the gene travel together on one or both alleles in these individuals, the individuals can be characterized as having a cytosine/cytosine haplotype with respect to the two SNPs in the gene.

[0037] The genetic analysis identified haplotypes associated with lower risk of fat deposition. In particular, presence of a haplotype represented by TTAT or TCAT at positions 11030, 15847, 21708, and 22713, respectively, in the P2X4 sequence represented by SEQ ID NO: 1 were associated with low central fat deposition. As used herein, a "haplotype" refers to a combination of polymorphic variations in a defined region within a genetic locus on one of the chromosomes in a chromosome pair.

[0038] In the genetic analysis that associated a polymorphic variation in P2X4 with NIDDM, samples from individuals in a population of diabetics and non-diabetics were genotyped. NIDDM can be diagnosed in patients by measuring fasting plasma glucose levels and/or casual plasma glucose levels, measuring fasting plasma insulin levels and/or casual plasma insulin levels, or administering oral glucose tolerance tests or hyperinsulinemic euglycemic clamp tests. A SNP at position 15847 in the P2X4 nucleotide sequence represented by SEQ ID NO:1 was genotyped in the diabetes case and control populations. It was determined that 33% of the individuals tested in the genetic analysis had a cytosine at position 15847 and 64% of the indivduals had a thymine at this position. It was determined that in females a thymine at position 15847 of SEQ ID NO:1 was individually associated with NIDDM, and in females the presence of a cytosine at position 15847 was individually associated with a not having NIDDM.

[0039] Additional Polymorphic Variants Associated with Fat Deposition

[0040] Also provided is a method for identifying polymorphic variants proximal to an incident, founder polymor-

phic variant associated with fat deposition, obesity and NIDDM. Thus, featured herein are methods for identifying a polymorphic variation associated with fat deposition or NIDDM that is proximal to an incident polymorphic variation associated with fat deposition or NIDDM, which comprises identifying a polymorphic variant proximal to the incident polymorphic variant associated with fat deposition of NIDDM, where the incident polymorphic variant is in a P2X4 nucleotide sequence. The P2X4 nucleotide sequence often comprises a polynucleotide sequence selected from the group consisting of (a) a polynucleotide sequence set forth in SEQ ID NO: 1; (b) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence encoded by a nucleotide sequence set forth as SEQ ID NO: 1; or (c) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence that is 90% identical to an amino acid sequence encoded by a nucleotide sequence set forth in SEQ ID NO: 1 or a polynucleotide sequence 90% identical to the polynucleotide sequence of SEQ ID NO:1. The presence or absence of an association of the proximal polymorphic variant with fat deposition or NIDDM then is determined using a known association method, such as a method described in the Examples hereafter. In an embodiment, the incident polymorphic variant is at position 11030, 15847, or 17338 of SEQ ID NO: 1. In another embodiment, the proximal polymorphic variant identified sometimes is a publicly disclosed polymorphic variant, which for example, sometimes is published in a publicly available database. In other embodiments, the polymorphic variant identified is not publicly disclosed and is discovered using a known method, including, but not limited to, sequencing a region surrounding the incident polymorphic variant in a group of nucleic samples. Thus, multiple polymorphic variants proximal to an incident polymorphic variant are associated with fat deposition and NIDDM using this method.

[0041] The proximal polymorphic variant often is identified in a region surrounding the incident polymorphic variant. In certain embodiments, this surrounding region is about 50 kb flanking the first polymorphic variant (e.g. about 50 kb 5' of the first polymorphic variant and about 50 kb 3' of the first polymorphic variant), and the region sometimes is composed of shorter flanking sequences, such as flanking sequences of about 40 kb, about 30 kb, about 25 kb, about 20 kb, about 15 kb, about 10 kb, about 7 kb, about 5 kb, or about 2 kb 5' and 3' of the incident polymorphic variant. In other embodiments, the region is composed of longer flanking sequences, such as flanking sequences of about 55 kb, about 60 kb, about 65 kb, about 70 kb, about 75 kb, about 80 kb, about 85 kb, about 90 kb, about 95 kb, or about 100 kb 5' and 3' of the incident polymorphic variant.

[0042] In certain embodiments, polymorphic variants associated with fat deposition or NIDDM are identified iteratively. For example, a first proximal polymorphic variant is associated with fat deposition using the methods described above and then another polymorphic variant proximal to the first proximal polymorphic variant is identified (e.g., publicly disclosed or discovered) and the presence or absence of an association of one or more other polymorphic variants proximal to the first proximal polymorphic variant with fat deposition or NIDDM is determined.

[0043] The methods described herein are useful for identifying or discovering additional polymorphic variants that may be used to further characterize a gene, region or loci associated with a condition, a disease (e.g., fat deposition or NIDDM), or a disorder. For example, allelotyping or genotyping data from the additional polymorphic variants may be used to identify a functional mutation or a region of linkage disequilibrium.

[0044] In certain embodiments, polymorphic variants identified or discovered within a region comprising the first polymorphic variant associated with fat deposition or NIDDM are genotyped using the genetic methods and sample selection techniques described herein (e.g., Example 6), and it can be determined whether those polymorphic variants are in linkage disequilibrium with the first polymorphic variant. The size of the region in linkage disequilibrium with the first polymorphic variant also can be assessed using these genotyping methods, and sizes of the regions surrounding an incident SNP described above often are probed for this determination. Thus, provided herein are methods for determining whether a proximal polymorphic variant is in linkage disequilibrium with an incident polymorphic variant associated with fat deposition or NIDDM, and such information is useful for the prognosis methods described herein. Example 6 and **FIG. 6** demonstrate that polymorphic variants in a region spanning from about position 11000 to about position 25300 in SEQ ID NO:1 or a substantially identical nucleic acid are in linkage disequilibrium. Polymorphic variants occur at positions 11030, 14744, 14977, 15046, 15059, 15847, 17338, 21538, 21630, 21708, 22713, and 25264, and other polymorphic variants in linkage disequilibrium may occur within the region or outside the region. Polymorphic variants occurring outside of the region often are within about 7 kb of position 11000 and position 25300, and sometimes within about 6 kb, about 5 kb, about 4 kb, about 3 kb, about 2 kb and about 1 kb of these positions.

[0045] Isolated P2X4 Nucleic Acids and Variants Thereof

[0046] Feartured herein are isolated P2X4 nucleic acids, which include the nucleic acid having the nucleotide sequence of SEQ ID NO:1, P2X4 nucleic acid variants, and substantially identical nucleic acids to the foregoing. Nucleotide sequences of the P2X4 nucleic acids are sometimes referred to herein as "P2X4 nucleotide sequences." A "P2X4 nucleic acid variant" refers to one allele that may have different polymorphic variations as compared to another allele in another subject or the same subject. A polymorphic variation in the P2X4 nucleic acid variant may be represented on one or both strands in a double-stranded nucleic acid or on one chromosomal complement (heterozygous) or both chromosomal complements (homozygous)). In certain embodiments, the nucleic acid comprises a P2X4 nucleotide sequence selected from the group consisting of: (a) the nucleotide sequence of SEQ ID NO:1; (b) a nucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; (c) a nucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3 or a nucleotide sequence that is 90% identical to the nucleotide sequence of SEQ ID NO:1; (d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and (e) a nucleotide sequence complementary to the nucleotide sequences of (a), (b), (c), or (d); where the P2X4 nucleotide sequence often comprises one or more polymorphic variations selected from

the group consisting of a thymine at position 14744, an adenine at position 14977, a cytosine at position 15046, a thymine at position 21538, a guanine at position 21630, and a single nucleotide deletion at position 25264 of SEQ ID NO:1, and optionally, a GGGCCCC nucleic acid insert between positions 2878 and 2891 in SEQ ID NO:1. The P2X4 polynucleotide sequence also sometimes comprises one or more polymorphic variations selected from the group consisting of an adenine at position 1075, an adenine at position 2383, a guanine at position 2494, a single nucleotide insert at position 2878, a thymine at position 3191, a guanine at position 25346, a cytosine at position 25798, a guanine at position 25864, an adenine at position 26334, a thymine at position 26343, and a guanine at position 27043 of SEQ ID NO:1. In specific embodiments, a P2X4 nucleic acid comprises one or more of the following polymorphic variations: an adenine or thymine at position 11030 of SEQ ID NO:1 in a strand, or a guanine or adenine in a complementary strand; a cytosine or thymine at position 15847 of SEQ ID NO:1 in a strand, or a guanine or adenine in a complementary strand; a cytosine or thymine at position 17338 of SEQ ID NO:1 in a strand, or a guanine or adenine in a complementary strand; presence of TTAT, TCAT, ACAT, or ACGC at positions 11030, 15847, 21708, and 22713 of SEQ ID NO:1, respectively, in a strand, or presence of AATA, AGTA, TGTA, or TGCG in a complementary strand; and presence or absence of a GGGCCCC insert between positions 2878 and 2891 of SEQ ID NO:1 in a strand, or presence or absence of a GGGGCCC insert in a complementary strand.

[0047] As used herein, the term "nucleic acid" includes DNA molecules (e.g., a complementary DNA (cDNA) and genomic DNA (gDNA)) and RNA molecules (e.g., mRNA, rRNA, siRNA and tRNA) and analogs of DNA or RNA, for example, by use of nucleotide analogs. The nucleic acid molecule can be single-stranded and it often is double-stranded. The term "isolated or purified nucleic acid" refers to nucleic acids that are separated from other nucleic acids present in the natural source of the nucleic acid. For example, with regard to genomic DNA, the term "isolated" includes nucleic acids which are separated from the chromosome with which the genomic DNA is naturally associated. An "isolated" nucleic acid often is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of 5' and/or 3' nucleotide sequences which flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. As used herein, the term "P2X4 gene" refers to a nucleotide sequence that encodes a P2X4 polypeptide.

[0048] Also included herein are nucleic acid fragments. These fragments are typically a nucleotide sequence identical to a nucleotide sequence in SEQ ID NO:1, a nucleotide sequence substantially identical to a nucleotide sequence in SEQ ID NO:1, or a nucleotide sequence that is complementary to the foregoing. The nucleic acid fragment may be identical, substantially identical or homolgous to a nucleotide sequence in an exon or an intron in SEQ ID NO:1 and may encode a domain or part of a domain of a P2X4 polypeptide. Sometimes, the fragment comprises one or more of the polymorphic variations described herein as being associated with increased fat deposition or increased risk of developing NIDDM. The nucleic acid fragment often is 50, 100, or 200 or fewer base pairs in length, and sometimes is about 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 2000, 3000, 4000, 5000, 10000, 15000, or 20000 base pairs in length. A nucleic acid fragment that is complementary to a nucleotide sequence identical or substantially identical to the nucleotide sequence of SEQ ID NO:1 and hybridizes to such a nucleotide sequence under stringent conditions often is referred to as a "probe." Nucleic acid fragments often include one or more polymorphic sites, or sometimes have an end that is adjacent to a polymorphic site as described hereafter.

[0049] An example of a nucleic acid fragment is an oligonucleotide. As used herein, the term "oligonucleotide" refers to a nucleic acid comprising about 8 to about 50 covalently linked nucleotides, often comprising from about 8 to about 35 nucleotides, and more often from about 10 to about 25 nucleotides. The backbone and nucleotides within an oligonucleotide may be the same as those of naturally occurring nucleic acids, or analogs or derivatives of naturally occurring nucleic acids, provided that oligonucleotides having such analogs or derivatives retain the ability to hybridize specifically to a nucleic acid comprising a targeted polymorphism. Oligonucleotides described hereafter are used as hybridization probes or as components of diagnostic assays, for example, as described herein.

[0050] Oligonucleotides typically are synthesized using standard methods and equipment, such as the ABI™3900 High Throughput DNA Synthesizer and the EXPEDITE™ 8909 Nucleic Acid Synthesizer, both of which are available from Applied Biosystems (Foster City, Calif.). Analogs and derivatives are exemplified in U.S. Pat. Nos. 4,469,863; 5,536,821; 5,541,306; 5,637,683; 5,637,684; 5,700,922; 5,717,083; 5,719,262; 5,739,308; 5,773,601; 5,886,165; 5,929,226; 5,977,296; 6,140,482; WO 00/56746; WO 01/14398, and related publications. Methods for synthesizing oligonucleotides comprising such analogs or derivatives are disclosed, for example, in the patent publications cited above and in U.S. Pat. Nos. 5,614,622; 5,739,314; 5,955,599; 5,962,674; 6,117,992; in WO 00/75372; and in related publications.

[0051] Oligonucleotides also may be linked to a second moiety. The second moiety may be an additional nucleotide sequence such as a tail sequence (e.g., a polyadenosine tail), an adaptor sequence (e.g., phage M13 universal tail sequence), and others. Alternatively, the second moiety may be a non-nucleotide moiety such as a moiety which facilitates linkage to a solid support or a label to facilitate detection of the oligonucleotide. Such labels include, without limitation, a radioactive label, a fluorescent label, a chemiluminescent label, a paramagnetic label, and the like. The second moiety may be attached to any position of the oligonucleotide, provided the oligonucleotide can hybridize to the nucleic acid comprising the polymorphism.

[0052] Uses for Nucleic Acids

[0053] Nucleic acid coding sequences depicted in SEQ ID NO:1, or substantially identical sequences thereof, may be

used for diagnostic purposes for detection and control of polypeptide expression. Also, included herein are oligonucleotide sequences such as antisense RNA, small-interfering RNA (siRNA) and DNA molecules and ribozymes that function to inhibit translation of a polypeptide. Antisense techniques and RNA interference techniques are known in the art and are described herein.

[0054] Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence specific hybridization of the ribozyme molecule to complementary target RNA, followed by a endonucleolytic cleavage. Ribozymes may be engineered hammerhead motif ribozyme molecules that specifically and efficiently catalyze endonucleolytic cleavage of RNA sequences corresponding to or complementary to the nucleotide sequences set forth in SEQ ID NO: 1. Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between fifteen (15) and twenty (20) ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for predicted structural features such as secondary structure that may render the oligonucleotide sequence unsuitable. The suitability of candidate targets may also be evaluated by testing their accessibility to hybridization with complementary oligonucleotides, using ribonuclease protection assays.

[0055] Antisense RNA and DNA molecules, siRNA and ribozymes are prepared by any method known in the art for the synthesis of nucleic acid molecules. These include known techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. RNA molecules sometimes are generated by in vitro and in vivo transcription of DNA sequences encoding the antisense RNA molecule. DNA sequences sometimes are incorporated into a vector which include a suitable RNA polymerase promoter such as the T7 or SP6 polymerase promoters. Also, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly, depending on the promoter used, sometimes are introduced into cell lines by stable transfection.

[0056] DNA encoding a polypeptide also is utilized in a number of methods, which include diagnosis of conditions or diseases, such as fat deposition or NIDDM, resulting from aberrant expression of the purinergic receptor. For example, the nucleic acid sequence may be used in hybridization assays of biopsies or autopsies to diagnose abnormalities of expression or function (e.g., Southern or Northern blot analysis, in situ hybridization assays).

[0057] In addition, the expression of a polypeptide during embryonic development may also be determined using nucleic acid encoding the polypeptide. As addressed, infra, production of functionally impaired polypeptide is the cause of various disease states, including fat deposition or NIDDM. In situ hybridizations using polypeptide as a probe may be employed to predict problems related to obesity or NIDDM. Further, as indicated, infra, administration of human active polypeptide, recombinantly produced as described herein, may be used to treat disease states related to functionally impaired polypeptide. Alternatively, gene therapy approaches may be employed to remedy deficiencies of functional polypeptide or to replace or compete with dysfunctional polypeptide.

[0058] Expression Vectors, Host Cells, and Genetically Engineered Cells

[0059] Provided herein are nucleic acid vectors, often expression vectors, which contain a P2X4 nucleic acid. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid, or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors may include replication defective retroviruses, adenoviruses and adeno-associated viruses for example.

[0060] A vector can include a P2X4 nucleic acid in a form suitable for expression of the nucleic acid in a host cell. The recombinant expression vector typically includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of polypeptide desired, and the like. Expression vectors can be introduced into host cells to produce P2X4 polypeptides, including fusion polypeptides, encoded by P2X4 nucleic acids.

[0061] Recombinant expression vectors can be designed for expression of P2X4 polypeptides in prokaryotic or eukaryotic cells. For example, P2X4 polypeptides can be expressed in *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

[0062] Expression of polypeptides in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion polypeptides. Fusion vectors add a number of amino acids to a polypeptide encoded therein, usually to the amino terminus of the recombinant polypeptide. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant polypeptide; 2) to increase the solubility of the recombinant polypeptide; and 3) to aid in the purification of the recombinant polypeptide by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant polypeptide to enable separation of the recombinant polypeptide from the fusion moiety subsequent to purification of the fusion polypeptide. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D. B. and Johnson, K. S., *Gene* 67: 31-40 (1988)), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) which fuse glutathione

S-transferase (GST), maltose E binding polypeptide, or polypeptide A, respectively, to the target recombinant polypeptide.

[0063] Purified fusion polypeptides can be used in screening assays and to generate antibodies specific for P2X4 polypeptides. In a therapeutic embodiment, fusion polypeptide expressed in a retroviral expression vector is used to infect bone marrow cells that are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six (6) weeks).

[0064] Expressing the polypeptide in host bacteria with an impaired capacity to proteolytically cleave the recombinant polypeptide often is used to maximize recombinant polypeptide expression (Gottesman, S., *Gene Expression Technology: Methods in Enzymology, Academic Press, San Diego, Calif.* 185: 119-128 (1990)). Another strategy is to alter the nucleotide sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al., Nucleic Acids Res. 20: 2111-2118 (1992)). Such alteration of nucleotide sequences can be carried out by standard DNA synthesis techniques.

[0065] When used in mammalian cells, the expression vector's control functions often are provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. Recombinant mammalian expression vectors often are capable of directing expression of the nucleic acid in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include an albumin promoter (liver-specific; Pinkert et al., *Genes Dev.* 1: 268-277 (1987)), lymphoid-specific promoters (Calame and Eaton, *Adv. Immunol.* 43: 235-275 (1988)), promoters of T cell receptors (Winoto and Baltimore, *EMBO J.* 8: 729-733 (1989)) promoters of immunoglobulins (Banerji et al., *Cell* 33: 729-740 (1983); Queen and Baltimore, *Cell* 33: 741-748 (1983)), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, *Proc. Natl. Acad. Sci. USA* 86: 5473-5477 (1989)), pancreas-specific promoters (Edlund et al., *Science* 230: 912-916 (1985)), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are sometimes utilized, for example, the murine hox promoters (Kessel and Gruss, *Science* 249: 374-379 (1990)) and the α-fetopolypeptide promoter (Campes and Tilghman, *Genes Dev.* 3: 537-546 (1989)).

[0066] A P2X4 nucleic acid may also be cloned into an expression vector in an antisense orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a P2X4 nucleic acid cloned in the antisense orientation can be chosen for directing constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. Antisense expression vectors can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H., et al., Antisense RNA as a molecular tool for genetic analysis, *Reviews—Trends in Genetics,* Vol. 1(1) (1986).

[0067] Also provided herein are host cells that include a P2X4 nucleic acid within a recombinant expression vector or P2X4 nucleic acid sequence fragments which allow it to homologously recombine into a specific site of the host cell genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but rather also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein. A host cell can be any prokaryotic or eukaryotic cell. For example, a P2X4 polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[0068] Vectors can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation.

[0069] A host cell provided herein can be used to produce (i.e., express) a P2X4 polypeptide. Accordingly, further provided are methods for producing a P2X4 polypeptide using the host cells of the invention. In one embodiment, the method includes culturing host cells into which a recombinant expression vector encoding a P2X4 polypeptide has been introduced in a suitable medium such that a P2X4 polypeptide is produced. In another embodiment, the method further includes isolating a P2X4 polypeptide from the medium or the host cell.

[0070] Also provided are cells or purified preparations of cells which include a P2X4 transgene, or which otherwise misexpress P2X4 polypeptide. Cell preparations can consist of human or non-human cells, e.g., rodent cells, e.g., mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a P2X4 transgene (e.g., a heterologous form of a P2X4 such as a human gene expressed in non-human cells). The P2X4 transgene can be misexpressed, e.g., overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene which misexpress an endogenous P2X4 polypeptide (e.g., expression of a gene is disrupted, also known as a knockout). Such cells can serve as a model for studying disorders which are related to mutated or mis-expressed P2X4 alleles or for use in drug screening. Also provided are human cells (e.g., a hematopoietic stem cells) transformed with a P2X4 nucleic acid.

[0071] Also provided are cells or a purified preparation thereof (e.g., human cells) in which an endogenous P2X4 nucleic acid is under the control of a regulatory sequence that does not normally control the expression of the endogenous P2X4 gene. The expression characteristics of an endogenous gene within a cell (e.g., a cell line or microorganism) can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the endogenous P2X4 gene. For example, an endogenous P2X4 gene (e.g., a gene which is "transcriptionally silent," not normally expressed, or expressed only at very low levels) may be

activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, e.g., Chappel, U.S. Pat. No. 5,272,071; WO 91/06667, published on May 16, 1991.

[0072] Transgenic Animals

[0073] Non-human transgenic animals that express a heterologous P2X4 polypeptide (e.g., expressed from a P2X4 nucleic acid isolated from another organism) can be generated. Such animals are useful for studying the function and/or activity of a P2X4 polypeptide and for identifying and/or evaluating modulators of P2X4 nucleic acid and P2X4 polypeptide activity. As used herein, a "transgenic animal" is a non-human animal such as a mammal (e.g., a non-human primate such as chimpanzee, baboon, or macaque; an ungulate such as an equine, bovine, or caprine; or a rodent such as a rat, a mouse, or an Israeli sand rat), a bird (e.g., a chicken or a turkey), an amphibian (e.g., a frog, salamander, or newt), or an insect (e.g., drosophila melanogaster), in which one or more of the cells of the animal includes a P2X4 transgene. A transgene is exogenous DNA or a rearrangement (e.g., a deletion of endogenous chromosomal DNA) that often is integrated into or occurs in the genome of cells in a transgenic animal. A transgene can direct expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, and other transgenes can reduce expression (e.g., a knockout). Thus, a transgenic animal can be one in which an endogenous P2X4 gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal (e.g., an embryonic cell of the animal) prior to development of the animal.

[0074] Intronic sequences and polyadenylation signals can also be included in the transgene to increase expression efficiency of the transgene. One or more tissue-specific regulatory sequences can be operably linked to a P2X4 transgene to direct expression of a P2X4 polypeptide to particular cells. A transgenic founder animal can be identified based upon the presence of a P2X4 transgene in its genome and/or expression of P2X4 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a P2X4 polypeptide can further be bred to other transgenic animals carrying other transgenes.

[0075] P2X4 polypeptides can be expressed in transgenic animals or plants by introducing, for example, a nucleic acid encoding the polypeptide into the genome of an animal. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, e.g., a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Also included is a population of cells from a transgenic animal.

[0076] P2X4 Polypeptides

[0077] Also featured herein are isolated P2X4 polypeptides, which include a polypeptide having the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3, P2X4 polypeptide variants, and substantially identical polypeptides thereof. A P2X4 polypeptide variant is a polypeptide encoded by a P2X4 nucleic acid variant. An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language "substantially free" means preparation of a P2X4 polypeptide or P2X4 polypeptide variant having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-P2X4 polypeptide (also referred to herein as a "contaminating protein"), or of chemical precursors or non-P2X4 chemicals. When the P2X4 polypeptide or a biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, specifically, where culture medium represents less than about 20%, sometimes less than about 10%, and often less than about 5% of the volume of the polypeptide preparation. Isolated or purified P2X4 polypeptide preparations are sometimes 0.01 milligrams or more or 0.1 milligrams or more, and often 1.0 milligrams or more and 10 milligrams or more in dry weight.

[0078] Further included herein are P2X4 polypeptide fragments. The polypeptide fragment may be a domain or part of a domain of a P2X4 polypeptide. P2X4 domains include, but are not limited to, an ATP P2X receptor domain at about amino acid positions 15 to 361 of SEQ ID NO:2. The polypeptide fragment may have increased, decreased or unexpected biological activity. The polypeptide fragment often is 50 or fewer, 100 or fewer, or 200 or fewer amino acids in length, and is sometimes 300 or 388 or fewer amino acids in length.

[0079] Substantially identical polypeptides may depart from the amino acid sequence of SEQ ID NO:2 in different manners. For example, conservative amino acid modifications may be introduced at one or more positions in the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3. A "conservative amino acid substitution" is one in which the amino acid is replaced by another amino acid having a similar structure and/or chemical function. Families of amino acid residues having similar structures and functions are well known. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Also, essential and non-essential amino acids may be replaced. A "non-essential" amino acid is one that can be altered without abolishing or substantially altering the biological function of a P2X4 polypeptide, whereas altering an "essential" amino acid abolishes or substantially alters the biological function of a P2X4 polypeptide. Amino acids that are conserved among purinergic receptors (e.g., P2X1, P2X2, P2X3, P2X4, P2X5, P2X6, and P2X7) are typically essential amino acids.

[0080] Also, P2X4 polypeptides and polypeptide variants may exist as chimeric or fusion polypeptides. As used herein, a P2X4 "chimeric polypeptide" or "fusion polypeptide" includes a P2X4 polypeptide linked to a non-P2X4 polypeptide. A "non-P2X4 polypeptide" refers to a polypep-

tide having an amino acid sequence corresponding to a polypeptide which is not substantially identical to the P2X4 polypeptide, which includes, for example, a polypeptide that is different from the P2X4 polypeptide and derived from the same or a different organism. The P2X4 polypeptide in the fusion polypeptide can correspond to an entire or nearly entire P2X4 polypeptide or a fragment thereof. The non-P2X4 polypeptide can be fused to the N-terminus or C-terminus of the P2X4 polypeptide.

[0081] Fusion polypeptides can include a moiety having high affinity for a ligand. For example, the fusion polypeptide can be a GST-P2X4 fusion polypeptide in which the P2X4 sequences are fused to the C-terminus of the GST sequences, or a polyhistidine-P2X4 fusion polypeptide in which the P2X4 polypeptide is fused at the N- or C-terminus to a string of histidine residues. Such fusion polypeptides can facilitate purification of recombinant P2X4. Expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide), and a P2X4 nucleic acid can be cloned into an expression vector such that the fusion moiety is linked in-frame to the P2X4 polypeptide. Further, the fusion polypeptide can be a P2X4 polypeptide containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression, secretion, cellular internalization, and cellular localization of a P2X4 polypeptide can be increased through use of a heterologous signal sequence. Fusion polypeptides can also include all or a part of a serum polypeptide (e.g., an IgG constant region or human serum albumin).

[0082] P2X4 polypeptides can be incorporated into pharmaceutical compositions and administered to a subject in vivo. Administration of these P2X4 polypeptides can be used to affect the bioavailability of a P2X4 substrate and may effectively increase P2X4 biological activity in a cell. P2X4 fusion polypeptides may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a P2X4 polypeptide; (ii) mis-regulation of the P2X4 gene; and (iii) aberrant post-translational modification of a P2X4 polypeptide. Also, P2X4 polypeptides can be used as immunogens to produce anti-P2X4 antibodies in a subject, to purify P2X4 ligands or binding partners, and in screening assays to identify molecules which inhibit or enhance the interaction of P2X4 with a P2X4 substrate. P2X4 polypeptides also sometimes are detected in a sample from a subject in a method for diagnosing or prognosing fat deposition or diabetes in the subject. In such embodiments, an alteration is detected in a P2X4 polypeptide that distinguishes it from P2X4 polypeptide in a subject not predisposed or diagnosed with fat deposition related disorders, such as NIDDM. Such alterations are detected in a variety of manners, which include, for example, contacting the sample with an antibody that specifically binds to a P2X4 polypeptide or altered P2X4 polypeptide, or sequencing regions of the polypeptide.

[0083] Substantially Identical P2X4 Nucleic Acids and Polypeptides

[0084] P2X4 nucleotide sequences and P2X4 polypeptide sequences that are substantially identical to the nucleotide sequence of SEQ ID NO:1 and the polypeptide sequence of SEQ ID NO:2. respectively, are included herein. The term "substantially identical" as used herein refers to two or more nucleic acids or polypeptides sharing one or more identical nucleotide sequences or polypeptide sequences, respectively. Included are nucleotide sequences or polypeptide sequences that are 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% (each often within a 1%, 2%, 3% or 4% variability) identical to the P2X4 nucleotide sequence in **FIG. 1** (SEQ ID NO:1) or the P2X4 polypeptide sequence of **FIG. 3A** (SEQ ID NO:2). One test for determining whether two nucleic acids are substantially identical is to determine the percent of identical nucleotide sequences or polypeptide sequences shared between the nucleic acids or polypeptides.

[0085] Calculations of sequence identity often are performed as follows. Sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is sometimes 30% or more, 40% or more, 50% or more, often 60% or more, and more often 70%, 80%, 90%, 100% of the length of the reference sequence. The nucleotides or amino acids at corresponding nucleotide or polypeptide positions, respectively, are then compared among the two sequences. When a position in the first sequence is occupied by the same nucleotide or amino acid as the corresponding position in the second sequence, the nucleotides or amino acids are deemed to be identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, introduced for optimal alignment of the two sequences.

[0086] Comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. Percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of E. Meyers and W. Miller, *CABIOS* 4: 11-17 (1989), which has been incorporated into the ALIGN program (version 2.0), using a PAM 120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. Also, percent identity between two amino acid sequences can be determined using the Needleman and Wunsch, *J. Mol. Biol.* 48: 444-453 (1970) algorithm which has been incorporated into the GAP program in the GCG software package (available at the http address www.gcg-.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. Percent identity between two nucleotide sequences can be determined using the GAP program in the GCG software package (available at http address www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A set of parameters often used is a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[0087] Another manner for determining if two nucleic acids are substantially identical is to assess whether a polynucleotide homologous to one nucleic acid will hybridize to the other nucleic acid under stringent conditions. As use herein, the term "stringent conditions" refers to conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y., 6.3.1-6.3.6 (1989). Aqueous and non-aqueous methods are described in that reference and either can be used. An

example of stringent hybridization conditions is hybridization in 6× sodium chloride/sodium citrate (SSC) at about 45° C., followed by one or more washes in 0.2×SSC, 0.1% SDS at 50° C. Another example of stringent hybridization conditions are hybridization in 6× sodium chloride/sodium citrate (SSC) at about 45° C., followed by one or more washes in 0.2×SSC, 0.1% SDS at 55° C. A further example of stringent hybridization conditions is hybridization in 6× sodium chloride/sodium citrate (SSC) at about 45° C., followed by one or more washes in 0.2×SSC, 0.1% SDS at 60° C. Often, stringent hybridization conditions are hybridization in 6× sodium chloride/sodium citrate (SSC) at about 45° C., followed by one or more washes in 0.2×SSC, 0.1% SDS at 65° C. More often, stringency conditions are 0.5M sodium phosphate, 7% SDS at 65° C., followed by one or more washes at 0.2×SSC, 1% SDS at 65° C.

[0088] An example of a substantially identical nucleotide sequence to SEQ ID NO:1 is one that has a different nucleotide sequence and still encodes the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:3. Another example is a nucleotide sequence that encodes a polypeptide having a polypeptide sequence that is more than 70% identical to, sometimes more than 75%, 80%, or 85% identical to, and often more than 90% and 95% identical to the polypeptide sequence of SEQ ID NO:2 or SEQ ID NO:3.

[0089] P2X4 nucleotide sequences and polypeptide sequences can be used as "query sequences" to perform a search against public databases to identify other family members or related sequences, for example. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul et al., *J. Mol. Biol.* 215: 403-10 (1990). BLAST nucleotide searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleotide sequences homologous to P2X4 nucleic acid molecules. BLAST polypeptide searches can be performed with the XBLAST program, score=50, wordlength=3 to obtain amino acid sequences homologous to P2X4 polypeptides. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., *Nucleic Acids Res.* 25(17): 3389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used (see the http address www.ncbi.nlm.nih.gov).

[0090] A nucleic acid that is substantially identical to the nucleotide sequence of SEQ ID NO:1 may include polymorphic sites at positions equivalent to those described herein (e.g., position 11030 in SEQ ID NO:1) when the sequences are aligned. For example, using the alignment procedures described herein, SNPs in a sequence substantially identical to the sequence of SEQ ID NO:1 can be identified at nucleotide positions that match (i.e., align) with nucleotides at SNP positions in SEQ ID NO:1. Also, where a polymorphic variation is an insertion or deletion, insertion or deletion of a nucleotide sequence from a reference sequence (e.g., a GGGCCCC insertion between positions 2878 to 2891 of SEQ ID NO:1) can change the relative positions of other polymorphic sites in the nucleotide sequence. For example, a GGGCCCC insertion within positions 2878 to 2891 in SEQ ID NO:1 will shift a SNP at position 11030 in SEQ ID NO:1 to position 11037. Also, a P2X4 nucleotide sequence sometimes includes an insert designated in **FIG. 5**, which can alter the nucleotide numbering designations after position 19137.

[0091] Substantially identical P2X4 nucleotide and polypeptide sequences include those that are naturally occurring, such as allelic variants (same locus), splice variants, homologs (different locus), and orthologs (different organism) or can be non-naturally occurring. Non-naturally occurring variants can be generated by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product). Orthologs, homologs, allelic variants, and splice variants can be identified using methods known in the art. These variants normally comprise a nucleotide sequence encoding a polypeptide that is 50%, about 55% or more, often about 70-75% or more, more often about 80-85% or more, and typically about 90-95% or more identical to the amino acid sequence shown in SEQ ID NO:2 or a fragment thereof. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions to the nucleotide sequence shown in SEQ ID NO:1 or a fragment of this sequence. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of the P2X4 nucleotide sequence can further be identified by mapping the sequence to the same chromosome or locus as the P2X4 nucleotide sequence or variant.

[0092] Also, substantially identical P2X4 nucleotide sequences may include codons that are altered with respect to the naturally occurring sequence for enhancing expression of a P2X4 polypeptide or polypeptide variant in a particular expression system. For example, the nucleic acid can be one in which one or more codons are altered, and often 10% or more or 20% or more of the codons are altered for optimized expression in bacteria (e.g., *E. coli.*), yeast (e.g., *S. cerevisiae*), human (e.g., 293 cells), insect, or rodent (e.g., hamster) cells.

[0093] Fat Deposition Disorder Prognostic and Diagnostic Methods

[0094] Methods for prognosing and diagnosing fat deposition, its related disorders (e.g., obesity and NIDDM) and leanness in subjects are provided herein. These methods include detecting the presence or absence of one or more polymorphic variations in a P2X4 nucleotide sequence or substantially identical sequence thereof in a sample from a subject, where the presence of a polymorphic variant described herein is indicative of a predisposition to leanness or fat deposition or one or more fat deposition related disorders (e.g., obesity or NIDDM). Determining a predisposition to fat deposition refers to determining whether an individual is at an increased or intermediate risk of fat deposition and determining a predisposition to leanness refers to a decreased risk of fat deposition. Determining a predisposition to NIDDM refers to determining whether an individual is at risk of NIDDM.

[0095] Thus, featured herein is a method for detecting a predisposition to fat deposition and a fat deposition disorder, such as obesity and NIDDM, in a subject, which comprises detecting the presence or absence of a polymorphic variation associated with fat deposition at a polymorphic site in a

P2X4 nucleotide sequence in a nucleic acid sample from a subject, wherein the P2X4 nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of: (a) the nucleotide sequence of SEQ ID NO:1; (b) a nucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; (c) a nucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3 or a nucleotide sequence about 90% or more identical to the nucleotide sequence of SEQ ID NO:1; and (d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising the polymorphic site; whereby the presence of the polymorphic variation is indicative of a predisposition to fat deposition in the subject. In certain embodiments, polymorphic variants at positions 11030, 15847 and 17338 are detected for determining a predisposition to fat deposition, a polymorphic variant at position 15847 is detected for determining a predisposition to NIDDM and polymorphic variants at positions in linkage disequilibrium with these positions are detected for determining a predisposition to fat deposition and NIDDM. Polymorphic variants within a region spanning from about position 11000 to about position 25300 in SEQ ID NO: 1 or a substantially identical nucleic acid sometimes are identified to assess whether a subject is predisposed to fat deposition and fat deposition disorders (e.g., obesity and NIDDM) as many polymorphic variants in this region are in linkage disequilibrium.

[0096] Results from prognostic tests may be combined with other test results to diagnose fat deposition related disorders, including NIDDM. For example, prognostic results may be gathered, a patient sample may be ordered based on a determined predisposition to fat deposition or NIDDM, the patient sample is analyzed, and the results of the analysis may be utilized to diagnose the fat deposition related condition (e.g., NIDDM). Also fat deposition diagnostic methods can be developed from studies used to generate prognostic methods in which populations are stratified into subpopulations having different progressions of a fat deposition related disorder or condition.

[0097] Predisposition to fat deposition, fat deposition related disorders such as NIDDM and obesity, and leanness sometimes is expressed as a probability, such as an odds ratio, percentage, or risk factor. The predisposition is based upon the presence or absence of one or more polymorphic variants described herein, and also may be based in part upon phenotypic traits of the individual being tested. Methods for calculating predispositions based upon patient data are well known (see, e.g., Agresti, *Categorical Data Analysis,* 2nd Ed. 2002. Wiley). Allelotyping and genotyping analyses may be carried out in populations other than those exemplified herein to enhance the predictive power of the prognostic method. These further analyses are executed in view of the exemplified procedures described herein, and may be based upon the same polymorphic variations or additional polymorphic variations.

[0098] The nucleic acid sample typically is isolated from a biological sample obtained from a subject. For example, nucleic acid can be isolated from blood, saliva, sputum, urine, cell scrapings, and biopsy tissue. The nucleic acid sample can be isolated from a biological sample using standard techniques, such as the technique described in Example 2. As used herein, the term "subject" refers primarily to humans but also refers to other mammals such as dogs, cats, and ungulates (e.g., cattle, sheep, and swine). Subjects also include avians (e.g., chickens and turkeys), reptiles, and fish (e.g., salmon), as embodiments described herein can be adapted to nucleic acid samples isolated from any of these organisms. The nucleic acid sample may be isolated from the subject and then directly utilized in a method for determining the presence of a polymorphic variant, or alternatively, the sample may be isolated and then stored (e.g., frozen) for a period of time before being subjected to analysis.

[0099] The presence or absence of a polymorphic variant is determined using one or both chromosomal complements represented in the nucleic acid sample. Determining the presence or absence of a polymorphic variant in both chromosomal complements represented in a nucleic acid sample from a subject having a copy of each chromosome is useful for determining the zygosity of an individual for the polymorphic variant (i.e., whether the individual is homozygous or heterozygous for the polymorphic variant). Any oligonucleotide-based diagnostic may be utilized to determine whether a sample includes the presence or absence of a polymorphic variant in a sample. For example, primer extension methods, ligase sequence determination methods (e.g., U.S. Pat. Nos. 5,679,524 and 5,952,174, and WO 01/27326), mismatch sequence determination methods (e.g., U.S. Pat. Nos. 5,851,770; 5,958,692; 6,110,684; and 6,183,958), microarray sequence determination methods, restriction fragment length polymorphism (RFLP), single strand conformation polymorphism detection (SSCP) (e.g., U.S. Pat. Nos. 5,891,625 and 6,013,499), PCR-based assays (e.g., TAQMAN® PCR System (Applied Biosystems)), and nucleotide sequencing methods may be used.

[0100] Oligonucleotide extension methods typically involve providing a pair of oligonucleotide primers in a polymerase chain reaction (PCR) or in other nucleic acid amplification methods for the purpose of amplifying a region from the nucleic acid sample that comprises the polymorphic variation. One oligonucleotide primer is complementary to a region 3' of the polymorphism and the other is complementary to a region 5' of the polymorphism. A PCR primer pair may be used in methods disclosed in U.S. Pat. Nos. 4,683,195; 4,683,202, 4,965,188; 5,656,493; 5,998,143; 6,140,054; WO 01/27327; and WO 01/27329 for example. PCR primer pairs may also be used in any commercially available machines that perform PCR, such as any of the GENEAMP® Systems available from Applied Biosystems. Also, those of ordinary skill in the art will be able to design oligonucleotide primers based upon the nucleotide sequence of SEQ ID NO:1 without undue experimentation using knowledge readily available in the art.

[0101] Also provided is an extension oligonucleotide that hybridizes to the amplified fragment adjacent to the polymorphic variation. As used herein, the term "adjacent" refers to the 3' end of the extension oligonucleotide being often 1 nucleotide from the 5' end of the polymorphic site, and sometimes 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides from the 5' end of the polymorphic site, in the nucleic acid when the extension oligonucleotide is hybridized to the nucleic acid. The extension oligonucleotide then is extended by one or more nucleotides, and the number and/or type of nucleotides that are added to the extension oligonucleotide determine whether the polymorphic variant is present. Oligonucleotide

extension methods are disclosed, for example, in U.S. Pat. Nos. 4,656,127; 4,851,331; 5,679,524; 5,834,189; 5,876,934; 5,908,755; 5,912,118; 5,976,802; 5,981,186; 6,004,744; 6,013,431; 6,017,702; 6,046,005; 6,087,095; 6,210,891; and WO 01/20039. Oligonucleotide extension methods using mass spectrometry are described, for example, in U.S. Pat. Nos. 5,547,835; 5,605,798; 5,691,141; 5,849,542; 5,869,242; 5,928,906; 6,043,031; and 6,194,144, and a method often utilized is described herein in Example 2.

[0102] A microarray can be utilized for determining whether a polymorphic variant is present or absent in a nucleic acid sample. A microarray may include any oligonucleotides described herein, and methods for making and using oligonucleotide microarrays suitable for diagnostic use are disclosed in U.S. Pat. Nos. 5,492,806; 5,525,464; 5,589,330; 5,695,940; 5,849,483; 6,018,041; 6,045,996; 6,136,541; 6,142,681; 6,156,501; 6,197,506; 6,223,127; 6,225,625; 6,229,911; 6,239,273; WO 00/52625; WO 01/25485; and WO 01/29259. The microarray typically comprises a solid support and the oligonucleotides may be linked to this solid support by covalent bonds or by non-covalent interactions. The oligonucleotides may also be linked to the solid support directly or by a spacer molecule. A microarray may comprise one or more oligonucleotides complementary to a polymorphic site of SEQ ID NO:1 (e.g., positions 11030, 15847 and 17338).

[0103] A kit also may be utilized for determining whether a polymorphic variant is present or absent in a nucleic acid sample. A kit often comprises one or more pairs of oligonucleotide primers useful for amplifying a fragment of SEQ ID NO:1 or a substantially identical sequence thereof, where the fragment includes a polymorphic site. The kit sometimes comprises a polymerizing agent, for example, a thermostable nucleic acid polymerase such as one disclosed in U.S. Pat. Nos. 4,889,818 or 6,077,664. Also, the kit often comprises an elongation oligonucleotide that hybridizes to a P2X4 nucleic acid in a nucleic acid sample adjacent to the polymorphic site. Where the kit includes an elongation oligonucleotide, it also often comprises chain elongating nucleotides, such as dATP, dTTP, dGTP, dCTP, and dITP, including analogs of dATP, dTTP, dGTP, dCTP and dITP, provided that such analogs are substrates for a thermostable nucleic acid polymerase and can be incorporated into a nucleic acid chain elongated from the extension oligonucleotide. Along with chain elongating nucleotides would be one or more chain terminating nucleotides such as ddATP, ddTTP, ddGTP, ddCTP, and the like. In an embodiment, the kit comprises one or more oligonucleotide primer pairs, a polymerizing agent, chain elongating nucleotides, at least one elongation oligonucleotide, and one or more chain terminating nucleotides. Kits optionally include buffers, vials, microtiter plates, and instructions for use.

[0104] Determining the presence of a polymorphic variant, or a combination of two or more polymorphic variants, in a P2X4 nucleic acid of the sample often is indicative of a predisposition to fat deposition or leanness. For example, presence of an adenine at position 11030 of SEQ ID NO:1 in the sense strand of a P2X4 nucleotide sequence is associated with an increased risk of fat deposition and presence of a thymine at position 11030 of SEQ ID NO:1 in the sense strand of a P2X4 nucleotide sequence is associated with leanness or a decreased risk of fat deposition. Specifically, a subject homozygous for an adenine at position 11030 of SEQ ID NO:1 in the sense strands of the P2X4 nucleotide sequence is at a higher risk of fat deposition, a subject heterozygous for an adenine and thymine at position 11030 in the sense strands of the P2X4 nucleotide sequence is at an intermediate risk of increased fat deposition, and a subject homozygous for a thymine at position 11030 in the sense strands of the P2X4 nucleotide sequence is at a lower risk of fat deposition. Similarly, a subject homozygous for a guanine at position 11030 in the strands complementary to the sense strands of the P2X4 nucleotide sequence is at a higher risk of increased fat deposition, a subject heterozygous for a guanine and adenine at position 11030 in the strands complementary to the sense strands of the P2X4 nucleotide sequence is at an intermediate risk of increased fat deposition, and a subject homozygous for an adenine at position 11030 in the strands complementary to the sense strands of the P2X4 nucleotide sequence is at a decreased risk of fat deposition.

[0105] Also, the presence of a thymine at position 15847 of SEQ ID NO:1 in the sense strand of a P2X4 nucleotide sequence is associated with an increased risk of fat deposition and NIDDM and the presence of a cytosine at position 15847 in the sense strand of a P2X4 nucleotide sequence is associated with leanness or a decreased risk of fat deposition and NIDDM. Specifically, a subject homozygous for a thymine at position 15847 of SEQ ID NO:1 in the sense strands of the P2X4 nucleotide sequence is at a higher risk of increased fat deposition and NIDDM, a subject heterozygous for a thymine and cytosine at position 15847 in the sense strands of the P2X4 nucleotide sequence is at an intermediate risk of increased fat deposition and NIDDM, and a subject homozygous for a cytosine at position 15847 in the sense strands of the P2X4 nucleotide sequence is at a decreased risk of fat deposition and NIDDM. Similarly, a subject homozygous for an adenine at position 15847 in the strands complementary to the sense strands of the P2X4 nucleotide sequence is at a higher risk of increased fat deposition and NIDDM, a subject heterozygous for a guanine and adenine at position 15847 in the strands complementary to the sense strands of the P2X4 nucleotide sequence is at an intermediate risk of increased fat deposition and NIDDM, and a subject homozygous for an guanine at position 15847 in the strands complementary to the sense strands of the P2X4 nucleotide sequence is at a lower risk of fat deposition and NIDDM.

[0106] Also, the presence of a cytosine at position 17338 of SEQ ID NO:1 in the sense strand of a P2X4 nucleotide sequence is associated with an increased risk of fat deposition and the presence of a thymine at position 17338 in the sense strand of a P2X4 nucleotide sequence is associated with leanness or a decreased risk of fat deposition. Specifically, a subject homozygous for a cytosine at position 17338 of SEQ ID NO:1 in the sense strands of the P2X4 nucleotide sequence is at a higher risk of increased fat deposition, a subject heterozygous for a cytosine and thymine at position 17338 in the sense strands of the P2X4 nucleotide sequence is at an intermediate risk of increased fat deposition, and a subject homozygous for a thymine at position 17338 in the sense strands of the P2X4 nucleotide sequence is at a decreased risk of fat deposition. Similarly, a subject homozygous for a guanine at position 17338 in the strands complementary to the sense strands of the P2X4 nucleotide sequence is at a higher risk of increased fat deposition, a subject heterozygous for a guanine and adenine at position

17338 in the strands complementary to the sense strands of the P2X4 nucleotide sequence is at an intermediate risk of increased fat deposition, and a subject homozygous for an adenine at position 17338 in the strands complementary to the sense strands of the P2X4 nucleotide sequence is at a lower risk of fat deposition.

[0107] Also, the presence of a haplotypes of TTAT and TCAT at positions 11030, 15847, 21708, and 22713, respectively, in the sense strand of a P2X4 nucleotide sequence (SEQ ID NO: 1) are associated with leanness or a decreased risk of fat deposition. Similarly, the presence of a haplotype of AATA and AGTA at positions 11030, 15847, 21708, and 22713, respectively, in the strand complementary to the sense strand of a P2X4 nucleotide sequence are associated a decreased risk of fat deposition. Haplotypes of ACAT and ACGC at positions 11030, 15847, 21708, and 22713, respectively, in the sense strand of a P2X4 nucleotide sequence are associated with an increased risk of fat deposition, as well as the presence of the haplotypes TGTA and TGCG at positions 11030, 15847, 21708, and 22713, respectively, in the strand complementary to the sense strand of a P2X4 nucleotide sequence.

[0108] Applications of Prognostic and Diagnostic Results to Pharmacogenomic Methods

[0109] Pharmacogenomics is a discipline that involves tailoring a treatment for a subject according to the subject's genotype as a particular treatment regimen may exert a differential effect depending upon the subject's genotype. Based upon the outcome of a prognostic test described herein, a clinician or physician may target pertinent information and preventative or therapeutic treatments to a subject who would be benefited by the information or treatment and avoid directing such information and treatments to a subject who would not be benefited (e.g., the treatment has no therapeutic effect and/or the subject experiences adverse side effects).

[0110] For example, where a candidate therapeutic exhibits a significant interaction with a major allele and a comparatively weak interaction with a minor allele (e.g. an order of magnitude or greater difference in the interaction), such a therapeutic typically would not be administered to a subject genotyped as being homozygous for the minor allele, and sometimes not administered to a subject genotyped as being heterozygous for the minor allele. In another example, where a candidate therapeutic is not significantly toxic when administered to subjects who are homozygous for a major allele but is comparatively toxic when administered to subjects heterozygous or homozygous for a minor allele, the candidate therapeutic is not typically administered to subjects who are genotyped as being heterozygous or homozygous with respect to the minor allele.

[0111] The prognostic methods described herein are applicable to pharmacogenomic methods for preventing, alleviating or treating fat deposition conditions such as obesity and NIDDM. For example, a nucleic acid sample from an individual may be subjected to a prognostic test described herein. Where one or more polymorphic variations associated with increased risk of obesity or NIDDM are identified in a subject, information for preventing or treating obesity or NIDDM and/or one or more obesity or NIDDM treatment regimens then may be prescribed to that subject. For example, a patient having a thymine at position 15847 in SEQ ID NO:1 often is prescribed a preventative regimen designed to minimize the occurrence of obesity or NIDDM.

[0112] In certain embodiments, a treatment regimen is specifically prescribed and/or administered to individuals who will most benefit from it based upon their risk of developing obesity or NIDDM assessed by the prognostic methods described herein. Thus, provided are methods for identifying a subject predisposed to obesity or NIDDM and then prescribing a therapeutic or preventative regimen to individuals identified as having a predisposition. Thus, certain embodiments are directed to a method for reducing fat deposition, obesity or NIDDM in a subject, which comprises: detecting the presence or absence of a polymorphic variant associated with fat deposition, obesity or NIDDM in a P2X4 nucleotide sequence in a nucleic acid sample from a subject, where the P2X4 nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of: (a) the polynucleotide sequence of SEQ ID NO:1; (b) a polynucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; (c) a polynucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; and (d) a fragment of a polynucleotide sequence of (i), (ii), or (iii); and prescribing or administering a treatment regimen to a subject from whom the sample originated where the presence of a polymorphic variation associated with fat reduction is detected in the P2X4 nucleotide sequence. In these methods, predisposition results may be utilized in combination with other test results to diagnose fat deposition associated conditions, such as obesity, metabolic conditions (e.g., NIDDM) and cardiovascular conditions (e.g., myocardial infarction).

[0113] The treatment sometimes is preventative (e.g., is prescribed or administered to reduce the probability that a fat deposition associated condition arises or progresses), sometimes is therapeutic, and sometimes delays, alleviates or halts the progression of a fat deposition associated condition. Any known preventative or therapeutic treatment for alleviating or preventing the occurrence of a fat deposition associated disorder is prescribed and/or administered. For example, the treatment sometimes is or includes a drug that reduces fat deposition, including, for example, an appetite suppressant (e.g., Phentermine, Adipex, Bontril, Didrex, Ionamin, Meridia, Phendimetrazine, Tenuate, Sibutramine), a lipase inhibitor (e.g., Olistat), a phospholipase inhibitor, a P2X4 nucleic acid, a P2X4 polypeptide, and/or a molecule that interacts with a P2X4 nucleic acid or P2X4 polypeptide described hereafter. In another example, the treatment is or includes a physical exercise regimen, dietary counseling and/or a dietary regimen (e.g., a low fat diet and/or a diet where the subject eats during pre-scheduled intervals) optionally coupled with dietary counseling, psychological counseling and/or psychotherapy, and sometimes optionally coupled with prescription of a psychotherapeutic or psychoprophylactic (e.g., an antidepressant or anti-anxiety therapeutic). In other embodiments directed to diabetes management, a subject sometimes is prescribed a regimen for regularly monitoring blood glucose levels, dietary counseling, a dietary regimen for managing blood glucose levels, and/or a blood glucose altering drug regimen. Examples of blood glucose altering drug regimens are regular administration of insulin (e.g., injection, pump, inhaler spray, nasal spray, insulin patch, and insulin tablet), and administration of hypoglycemics (e.g., glyburide or repaglinide), starch

blockers (e.g., acarbose), liver glucose regulating agents (e.g., metformin), and/or insulin sensitizers (e.g., rosiglitazone or pioglitazone). Prescription and/or administration of each treatment or combinations of treatments often is dependent upon the age of the subject as well as the subject's physiological, medical, and/or psychological condition.

[0114] In an embodiment, the pharmacogenomic methods described herein are applicable to subjects who are women about forty or more years of age and have not yet entered menopause, undergoing menopause, or post-menopausal. Those subjects identified as having an increased risk for fat deposition sometimes are prescribed a hormone replacement treatment (HRT) regimen. There are many HRT regimens known in the art, which include regular administration of estrogen (e.g., Prumarin®), progesterone (e.g., Provera®), androgen (e.g., testosterone), a combination of estrogen and progesterone, a combination of estrogen and androgen (e.g., Estratest®), growth hormone, dehydroepiandrosterone (DHEA), a sulfate ester of DHEA, or a combination of DHEA and a DHEA sulfate ester. Also, selective estrogen receptor modulators (SERMs) such as raloxifene and tamoxifen, for example, can be prescribed. Those women diagnosed as having an increased risk of fat deposition sometimes are prescribed an estrogen replacement therapy (ERT) regimen or SERMs regimen as an alternative to a combination of estrogen and progesterone, due to an association between ERT and lower fat deposition and an association between increased fat deposition and progesterone replacement therapy.

[0115] In another embodiment, pharmacogenomic methods are applicable to subjects who are women using a contraceptive or are contemplating use of a contraceptive, where the contraceptive has been shown to increase fat deposition in subjects. This embodiment often applies to women who are pre-pubescent, who are in puberty, or who are post-pubescent and pre-menopausal. Many oral contraceptives, especially those that include higher contents of estrogen compared to other oral contraceptives, have been shown to increase fat deposition in subjects. Those subjects identified as having an increased risk for fat deposition by the methods described herein often are advised not to begin an oral contraceptive regimen or to discontinue an oral contraceptive regimen. Alternatively, subjects identified as having an increased risk for fat deposition sometimes are advised to begin an oral contraceptive regimen using a contraceptive having lower estrogen content as compared to other available oral contraceptives (e.g., Allesse®, Levlite®, Loestrin-Fe®, and Mircette® are examples of contraceptives having lower estrogen content).

[0116] As therapeutic approaches for obesity or NIDDM continue to evolve and improve, the goal of treatments for fat deposition related disorders is to intervene even before clinical signs (e.g., impaired glucose tolerance) first manifest. Thus, genetic markers associated with susceptibility to obesity or NIDDM prove useful for early diagnosis, prevention and treatment of obesity or NIDDM.

[0117] As obesity or NIDDM preventative and treatment information can be specifically targeted to subjects in need thereof (e.g. those at risk of developing obesity or NIDDM or those that have early stages of obesity or NIDDM), provided herein is a method for preventing or reducing the risk of developing obesity or NIDDM in a subject, which comprises: (a) detecting the presence or absence of a polymorphic variation associated with obesity or NIDDM at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) identifying a subject with a predisposition to obesity or NIDDM, whereby the presence of the polymorphic variation is indicative of a predisposition to obesity or NIDDM in the subject; and (c) if such a predisposition is identified, providing the subject with information about methods or products to prevent or reduce obesity or NIDDM or to delay the onset of obesity or NIDDM. Also provided is a method of targeting information or advertising to a subpopulation of a human population based on the subpopulation being genetically predisposed to a disease or condition, which comprises: (a) detecting the presence or absence of a polymorphic variation associated with obesity or NIDDM at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) identifying the subpopulation of subjects in which the polymorphic variation is associated with obesity or NIDDM; and (c) providing information only to the subpopulation of subjects about a particular product which may be obtained and consumed or applied by the subject to help prevent or delay onset of the disease or condition.

[0118] Pharmacogenomics methods also may be used to analyze and predict a response to an obesity or NIDDM treatment or a drug. For example, if pharmacogenomics analysis indicates a likelihood that an individual will respond positively to a obesity or NIDDM treatment with a particular drug, the drug may be administered to the individual. Conversely, if the analysis indicates that an individual is likely to respond negatively to treatment with a particular drug, an alternative course of treatment may be prescribed. A negative response may be defined as either the absence of an efficacious response or the presence of toxic side effects. The response to a therapeutic treatment can be predicted in a background study in which subjects in any of the following populations are genotyped: a population that responds favorably to a treatment regimen, a population that does not respond significantly to a treatment regimen, and a population that responds adversely to a treatment regiment (e.g. exhibits one or more side effects). These populations are provided as examples and other populations and subpopulations may be analyzed. Based upon the results of these analyses, a subject is genotyped to predict whether he or she will respond favorably to a treatment regimen, not respond significantly to a treatment regimen, or respond adversely to a treatment regimen.

[0119] The prognostic tests described herein also are applicable to clinical drug trials. One or more polymorphic variants indicative of response to an agent for treating obesity or NIDDM or to side effects to an agent for treating obesity or NIDDM may be identified using the methods described herein. Thereafter, potential participants in clinical trials of such an agent may be screened to identify those individuals most likely to respond favorably to the drug and exclude those likely to experience side effects. In that way, the effectiveness of drug treatment may be measured in individuals who respond positively to the drug, without lowering the measurement as a result of the inclusion of individuals who are unlikely to respond positively in the study and without risking undesirable safety problems.

[0120] Thus, another embodiment is a method of selecting an individual for inclusion in a clinical trial of a treatment or

drug comprising the steps of: a) obtaining a nucleic acid sample from an individual; b) determining the identity of a polymorphic variation which is associated with a positive response to the treatment or the drug, or at least one polymorphic variation which is associated with a negative response to the treatment or the drug in the nucleic acid sample, and c) including the individual in the clinical trial if the nucleic acid sample contains said polymorphic variation associated with a positive response to the treatment or the drug or if the nucleic acid sample lacks said polymorphic variation associated with a negative response to the treatment or the drug. In addition, the methods of the present invention for selecting an individual for inclusion in a clinical trial of a treatment or drug encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination. The polymorphic variation may be in a sequence selected individually or in any combination from the group consisting of (i) a polynucleotide sequence set forth in SEQ ID NO: 1; (ii) a polynucleotide sequence that is 90% identical to a nucleotide sequence set forth in SEQ ID NO: 1; (iii) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence identical to or 90% identical to an amino acid sequence encoded by a nucleotide sequence set forth in SEQ ID NO: 1; and (iv) a fragment of a polynucleotide sequence of (i), (ii), or (iii) comprising the polymorphic site. The including step (c) optionally comprises administering the drug or the treatment to the individual if the nucleic acid sample contains the polymorphic variation associated with a positive response to the treatment or the drug and the nucleic acid sample lacks said biallelic marker associated with a negative response to the treatment or the drug.

[0121] Also provided herein is a method of partnering between a diagnostic/prognostic testing provider and a provider of a consumable product, which comprises: (a) the diagnostic/prognostic testing provider detects the presence or absence of a polymorphic variation associated with obesity or NIDDM at a polymorphic site in a nucleotide sequence in a nucleic acid sample from a subject; (b) the diagnostic/prognostic testing provider identifies the subpopulation of subjects in which the polymorphic variation is associated with obesity or NIDDM; (c) the diagnostic/prognostic testing provider forwards information to the subpopulation of subjects about a particular product which may be obtained and consumed or applied by the subject to help prevent or delay onset of the disease or condition; and (d) the provider of a consumable product forwards to the diagnostic test provider a fee every time the diagnostic/prognostic test provider forwards information to the subject as set forth in step (c) above. Methods for Identifying Candidate Therapeutics for Reducing Fat Deposition

[0122] Current therapies for the treatment of NIDDM have limited efficacy, limited tolerability and significant mechanisms-based side effects, including weight gain and hypoglycemia. Few of the available therapies adequately address underlying defects such as obesity and insulin resistance. Thus, newer approaches are desperately needed (Moller D. Nature. 414:821-827 (2001)). Current therapeutic approaches were largely developed in the absence of defined molecular targets or even a solid understanding of disease pathogenesis. The same holds true for the treatment of obesity, where treatments have limited lasting effects and many side effects. Therefore, there is a need for methods of identifying candidate therapeutics that target the biochemical pathways related to the development of obesity and/or diabetes.

[0123] Featured herein is a method for identifying candidate therapeutics for reducing fat deposition, obesity and/or alleviating NIDDM. The method often comprises introducing a test molecule to a system which comprises a nucleic acid comprising a P2X4 nucleotide sequence selected from the group consisting of: the nucleotide sequence of SEQ ID NO:1; a nucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; a nucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3 or a nucleotide sequence 90% identical to the nucleotide sequence of SEQ ID NO:1; and a fragment of one of the foregoing nucleotide sequences. Alternatively, the method comprises introducing a test molecule to a system which comprises a protein encoded by one of the foregoing nucleotide sequences. The presence or absence of an interaction between the test molecule and the nucleic acid or protein is determined and the presence of an interaction between the test molecule and the nucleic acid or protein identifies the test molecule as a candidate therapeutic for fat reduction, obesity and/or alleviating NIDDM.

[0124] As used herein, the term "test molecule" and "candidate therapeutic" refers to modulators that regulate transcription and translation of P2X4 nucleic acids and modulators of expression and activity of P2X4 polypeptides. The term "modulator" as used herein refers to a molecule which agonizes or antagonizes P2X4 DNA replication and/or DNA processing (e.g., methylation), P2X4 RNA transcription and/or RNA processing (e.g., removal of intronic sequences and/or translocation from the nucleus), P2X4 polypeptide production (e.g., translation of the polypeptide from mRNA, and/or post-translational modification such as glycosylation, phosphorylation, and proteolysis of pro-polypeptides), and/or P2X4 function (e.g., conformational changes, binding of nucleotides or nucleotide analogs, binding and/or translocation of ions, interaction with binding partners, effect on membrane potential, effect on fat deposition, effect on metabolic condition, and/or effect on cardivascular condition). Test molecules and candidate therapeutics include, but are not limited to, compounds, antisense nucleic acids, siRNA molecules, ribozymes, P2X4 polypeptide or fragments thereof, immunotherapeutics (e.g., antibodies).

[0125] Compounds

[0126] Compounds may be utilized as test molecules for identifying candidate therapeutics for reducing fat deposition. Compounds can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive (see, e.g., Zuckermann, R. N. et al., *J. Med. Chem.* 37: 2678-85 (1994)); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; "one-bead one-compound" library methods; and synthetic library methods using affinity chromatography selection. Biological library and peptoid library approaches are typically limited to peptide libraries, while the other approaches are applicable to peptide, non-peptide

oligomer or small molecule libraries of compounds (Lam, *Anticancer Drug Des.* 12: 145, (1997)). Examples of methods for synthesizing molecular libraries are described, for example, in DeWitt et al., *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909 (1993); Erb et al., *Proc. Natl. Acad. Sci. USA* 91: 11422 (1994); Zuckermann et al., *J. Med. Chem.* 37: 2678 (1994); Cho et al., Science 261: 1303 (1993); Carrell et al., *Angew. Chem. Int. Ed. Engl.* 33. 2059 (1994); Carell et al., *Angew. Chem. Int. Ed. Engl.* 33: 2061 (1994); and in Gallop et al., *J. Med. Chem.* 37. 1233 (1994).

[0127] Libraries of compounds may be presented in solution (e.g., Houghten, *Biotechniques* 13; 412-421 (1992)), or on beads (Lam, *Nature* 354: 82-84 (1991)), chips (Fodor, *Nature* 364: 555-556 (1993)), bacteria or spores (Ladner, U.S. Pat. No. 5,223,409), plasmids (Cull et al., *Proc. Natl. Acad. Sci. USA* 89: 1865-1869 (1992)) or on phage (Scott and Smith, *Science* 249: 386-390 (1990); Devlin, *Science* 249: 404-406 (1990); Cwirla et al., *Proc. Natl. Acad. Sci.* 87: 6378-6382 (1990); Felici, *J. Mol. Biol.* 222: 301-310 (1991); Ladner supra.).

[0128] Compounds may alter expression or activity of P2X4 polypeptides and may be a small molecule. Small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

[0129] Antisense, Ribozyme. siRNA, and Modified P2X4 Nucleic Acid Molecules

[0130] Also featured herein are antisense, ribozyme, siRNA, and modified P2X4 nucleic acids for use as test molecules in methods for identifying candidate therapeutics for reducing fat deposition or treating NIDDM, and for use as therapeutics for reducing fat deposition or treating NIDDM in a subject. An "antisense" nucleic acid refers to a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a polypeptide, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire P2X4 coding strand, or to only a portion thereof (e.g., the coding region of human P2X4 corresponding to SEQ ID NO: 1). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding P2X4 (e.g., 5' and 3' untranslated regions).

[0131] An antisense nucleic acid can be designed such that it is complementary to the entire coding region of P2X4 mRNA, and often the antisense nucleic acid is an oligonucleotide that is antisense to only a portion of a coding or noncoding region of P2X4 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of P2X4 mRNA, e.g., between the −10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length. The antisense nucleic acids, which include the ribozymes described hereafter, can be designed to target P2X4 nucleic acid or P2X4 nucleic acid variants. Among the variants, minor alleles and major alleles can be targeted, and those associated with a higher risk to fat deposition, such as alleles having an adenine at position 11030, a thymine at position 15847 or a cytosine at position 17338 in the P2X4 nucleotide sequence represented by SEQ ID NO:1, often are designed, tested, and administered to subjects. Similarly, antisense nucleic acids, which include the ribozymes described hereafter, can be designed to target P2X4 nucleic acid or P2X4 nucleic acid variants associated with a higher risk of developing NIDDM, such as nucleic acids having a thymine at positions 15847 of SEQ ID NO:1. An antisense nucleic acid can be constructed using chemical synthesis and enzymatic ligation reactions using standard procedures. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[0132] Antisense nucleic acids are typically administered to a subject (e.g., by direct injection at a tissue site) or generated in situ such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a P2X4 polypeptide and thereby inhibit expression of the polypeptide, for example, by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, for example, by linking antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. Antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. Sufficient intracellular concentrations of antisense molecules are achieved by incorporating a strong promoter, such as a pol II or pol III promoter, in the vector construct.

[0133] Antisense nucleic acid molecules are sometimes α-anomeric nucleic acid molecules. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier et al., *Nucleic Acids. Res.* 15: 6625-6641 (1987)). Antisense nucleic acid molecules can also comprise a 2'-o-methylribonucleotide (Inoue et al., *Nucleic Acids Res.* 15: 6131-6148 (1987)) or a chimeric RNA-DNA analogue (Inoue et al., *FEBS Lett.* 215: 327-330 (1987)).

[0134] In another embodiment, an antisense nucleic acid is a ribozyme. A ribozyme having specificity for a P2X4-encoding nucleic acid can include one or more sequences

complementary to the nucleotide sequence of a P2X4 DNA sequence disclosed herein (e.g., SEQ ID NO:1), and a sequence having a known catalytic sequence responsible for mRNA cleavage (see U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach, *Nature* 334: 585-591 (1988)). For example, a derivative of a Tetrahymena L-19 IVS RNA is sometimes utilized in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a P2X4-encoding mRNA. See, e.g., Cech et al. U.S. Pat. No. 4,987,071; and Cech et al. U.S. Pat. No. 5,116,742. Also, P2X4 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel & Szostak, *Science* 261: 1411-1418 (1993).

[0135] P2X4 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the P2X4 (e.g., P2X4 promoter and/or enhancers) to form triple helical structures that prevent transcription of the P2X4 gene in target cells. See, Helene, *Anticancer Drug Des.* 6(6): 569-84 (1991); Helene et al., *Ann. N.Y. Acad. Sci.* 660: 27-36 (1992); and Maher, *Bioassays* 14(12): 807-15 (1992). Potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

[0136] P2X4 expression may be inhibited by the introduction of double-stranded RNA (dsRNA), which induces potent and specific gene silencing, a phenomenon called RNA interference or RNAi. See, e.g., Fire et al., U.S. Pat. No. 6,506,559; Tuschl et al. PCT International Publication No. WO 01/75164; Kay et al. PCT International Publication No. WO 03/010180A 1; or Bosher J M, Labouesse, *Nat Cell Biol February* 2000;2(2):E31-6. This process has been improved by decreasing the size of the double-stranded RNA to 20-24 base pairs (to create small-interfering RNAs or siRNAs) that "switched off" genes in mammalian cells without initiating an acute phase response, i.e., a host defense mechanism that often results in cell death. See, e.g., Caplen et al. *Proc Natl Acad Sci USA. Aug.* 14, 2001;98(17):9742-7 and Elbashir S M et al. *Methods February* 2002;26(2):199-213. There is increasing evidence that post-transcriptional gene silencing by RNA interference (RNAi) for inhibiting targeted expression in mammalian cells at the mRNA level is effective in-human cells. There is additional evidence of effective methods for inhibiting the proliferation and migration of tumor cells in human patients, and for inhibiting metastatic cancer development. See, e.g., U.S. Patent Application Number US2001000993183; Caplen N J et al. *Proc Natl Acad Sci* USA; and Abderrahmani A. et al. *Mol Cell Biol Nov.* 21, 2001(21):7256-67.

[0137] An "siRNA" or "RNAi" refers to a nucleic acid that forms a double stranded RNA and has the ability to reduce or inhibit expression of a gene or target gene when the siRNA is delivered to or expressed in the same cell as the gene or target gene. "siRNA" thus refers to short double stranded RNA formed by the complementary strands. Complementary portions of the siRNA that hybridize to form the double stranded molecule often have substantial or complete identity to the target molecule sequence. In one embodiment, an siRNA refers to a nucleic acid that has substantial or complete identity to a target gene and forms a double stranded siRNA, such as a nucleotide sequence in SEQ ID NO: 1, for example.

[0138] When designing the siRNA molecules, the targeted region often is selected from a given DNA sequence beginning 50 to 100 nt downstream of the start codon. See, e.g., Elbashir et al, *Methods* 26:199-213 (2002). Initially, 5' or 3' UTRs and regions nearby the start codon were avoided assuming that UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP or RISC endonuclease complex. Sometimes regions of the target 23 nucleotides in length conforming to the sequence motif AA(N19)TT (N, an nucleotide), and regions with approximately 30% to 70% G/C-content (often about 50% G/C-content) often are selected. If no suitable sequences are found, the search often is extended using the motif NA(N21). The sequence of the sense siRNA sometimes corresponds to (N19) TT or N21 (position 3 to 23 of the 23-nt motif), respectively. In the latter case, the 3' end of the sense siRNA often is converted to TT. The rationale for this sequence conversion is to generate a symmetric duplex with respect to the sequence composition of the sense and antisense 3' overhangs. The antisense siRNA is synthesized as the complement to position 1 to 21 of the 23-nt motif. Because position I of the 23-nt motif is not recognized sequence-specifically by the antisense siRNA, the 3'-most nucleotide residue of the antisense siRNA can be chosen deliberately. However, the penultimate nucleotide of the antisense siRNA (complementary to position 2 of the 23-nt motif) often is complementary to the targeted sequence. For simplifying chemical synthesis, TT often is utilized. siRNAs corresponding to the target motifNAR(N17)YNN, where R is purine (A,G) and Y is pyrimidine (C,U), often are selected. Respective 21 nucleotide sense and antisense siRNAs often begin with a purine nucleotide and can also be expressed from pol III expression vectors without a change in targeting site. Expression of RNAs from pol III promoters often is efficient when the first transcribed nucleotide is a purine.

[0139] The sequence of the siRNA can correspond to the full length target gene, or a subsequence thereof. Often, the siRNA is about 15 to about 50 nucleotides in length (e.g., each complementary sequence of the double stranded siRNA is 15-50 nucleotides in length, and the double stranded siRNA is about 15-50 base pairs-in length, sometimes about 20-30 nucleotides in length or about 20-25 nucleotides in length, e.g., 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length. The siRNA often is about 21 nucleotides in length. Methods of using siRNA are well known in the art, and specific siRNA molecules may be purchased from a number of companies including Dharmacon Research, Inc.

[0140] Antisense, ribozyme, and modified P2X4 nucleic acid molecules can be altered at base moieties, sugar moieties or phosphate backbone moieties to improve stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup et al., *Bioorganic & Medicinal Chemistry* 4 (1). 5-23 (1996)). As used herein, the terms "peptide nucleic acid" or "PNA" refers to a nucleic acid mimic such as a DNA mimic, in which the deoxyribose phosphate backbone is replaced by

a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. Synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described, for example, in Hyrup et al., (1996) supra and Perry-O'Keefe et al., *Proc. Natl. Acad. Sci.* 93: 14670-675 (1996).

[0141] PNAs of P2X4 nucleic acids can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of P2X4 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as "artificial restriction enzymes" when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup (1996) supra)); or as probes or primers for DNA sequencing or hybridization (Hyrup et al., (1996) supra; Perry-O'Keefe supra).

[0142] In other embodiments, oligonucleotides may include other appended groups such as peptides (e.g., for targeting host cell receptors in vivo), or agents facilitating transport across cell membranes (see, e.g., Letsinger et al., *Proc. Natl. Acad. Sci. USA* 86: 6553-6556 (1989); Lemaitre et al., *Proc. Natl. Acad. Sci. USA* 84: 648-652 (1987); PCT Publication No. WO88/09810) or the blood-brain barrier (see, e.g., PCT Publication No. WO89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, e.g., Krol et al., *Bio-Techniques* 6: 958-976 (1988)) or intercalating agents. (See, e.g., Zon, *Pharm. Res.* 5. 539-549 (1988)). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[0143] Also included herein are molecular beacon oligonucleotide primer and probe molecules having one or more regions which are complementary to a P2X4 nucleic acid of the invention, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantifying the presence of the P2X4 nucleic acid of the invention in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi et al., U.S. Pat. No. 5,854,033; Nazarenko et al., U.S. Pat. No. 5,866,336, and Livak et al., U.S. Pat. No. 5,876,930.

[0144] Anti-P2X4 Antibodies

[0145] In an embodiment, antibodies are screened as test molecules and used as therapeutics for reducing fat deposition or treating NIDDM in a subject. The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. An antibody can be a polyclonal, monoclonal, recombinant, e.g., a chimeric or humanized, fully human, non-human, e.g., murine, or single chain antibody. An antibody may have effector function and can fix complement, and is sometimes coupled to a toxin or imaging agent.

[0146] A full-length P2X4 polypeptide or, antigenic peptide fragment of P2X4 can be used as an immunogen or can be used to identify anti-P2X4 antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. The antigenic peptide of P2X4 should include at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of P2X4. Antigenic peptides include 10 or more amino acids, 15 or more amino acids, often 20 or more amino acids, and typically 30 or more amino acids. Hydrophilic and hydrophobic fragments of P2X4 polypeptides can be used as immunogens.

[0147] Epitopes encompassed by the antigenic peptide are regions of P2X4 located on the surface of the polypeptide (e.g., hydrophilic regions) as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human P2X4 polypeptide sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface of the P2X4 polypeptide and are thus likely to constitute surface residues useful for targeting antibody production. The antibody may bind an epitope on any domain or region on P2X4 polypeptides described herein.

[0148] Also, chimeric, humanized, and completely human antibodies are useful for applications which include repeated administration to subjects. Chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, can be made using standard recombinant DNA techniques. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson et al. International Application No. PCT/US86/02269; Akira, et al. European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al. European Patent Application 173,494; Neuberger et al. PCT International Publication No. WO 86/01533; Cabilly et al. U.S. Pat. No. 4,816,567; Cabilly et al. European Patent Application 125,023; Better et al., *Science* 240: 1041-1043 (1988); Liu et al., *Proc. Natl. Acad. Sci. USA* 84: 3439-3443 (1987); Liu et al., *J. Immunol.* 139: 3521-3526 (1987); Sun et al., *Proc. Natl. Acad. Sci. USA* 84: 214-218 (1987); Nishimura et al., *Canc. Res.* 47: 999-1005 (1987); Wood et al., *Nature* 314: 446-449 (1985); and Shaw et al., *J. Natl. Cancer Inst.* 80: 1553-1559 (1988); Morrison, S. L., *Science* 229: 1202-1207 (1985); Oi et al., BioTechniques 4: 214 (1986); Winter U.S. Pat. No. 5,225,539; Jones et al., *Nature* 321. 552-525 (1986); Verhoeyan et al., *Science* 239: 1534; and Beidler et al., *J. Immunol.* 141: 4053-4060 (1988).

[0149] Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice that are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. See, for example, Lonberg and Huszar, *Int. Rev. Immunol.* 13: 65-93 (1995); and U.S. Pat. Nos. 5,625,126; 5,633,425; 5,569,825; 5,661,016; and 5,545,806. In addition, companies such as Abgenix, Inc. (Fremont, Calif.) and Medarex, Inc. (Princeton, N.J.), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above. Completely human antibodies that recognize a selected epitope also can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody (e.g., a murine antibody) is used to guide the selection of a completely human antibody

recognizing the same epitope. This technology is described for example by Jespers et al., *Bio/Technology* 12. 899-903 (1994).

[0150] An anti-P2X4 antibody can be a single chain antibody. A single chain antibody (scFV) can be engineered (see, e.g., Colcher, D. et al., *Ann. NY Acad. Sci.* 880: 263-80 (1999); and Reiter, Y., *Clin. Cancer Res.* 2: 245-52 (1996)). Single chain antibodies can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target P2X4 polypeptide.

[0151] Antibodies also may be selected or modified so that they exhibit reduced or no ability to bind an Fc receptor. For example, an antibody may be an isotype or subtype, fragment or other mutant, which does not support binding to an Fc receptor (e.g., it has a mutagenized or deleted Fc receptor binding region).

[0152] Also, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine and vinblastine).

[0153] Antibody conjugates can be used for modifying a given biological response. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a polypeptide such as tumor necrosis factor, α-interferon, β-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors. Also, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Pat. No. 4,676,980, for example.

[0154] An anti-P2X4 antibody (e.g., monoclonal antibody) can be used to isolate P2X4 polypeptides by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an anti-P2X4 antibody can be used to detect a P2X4 polypeptide (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. Anti-P2X4 antibodies can be used diagnostically to monitor polypeptide levels in tissue as part of a clinical testing procedure, e.g., to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labeling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

[0155] An antibody can be made by immunizing with a purified P2X4 antigen, or a fragment thereof, e.g., a fragment described herein, a membrane associated antigen, tissues, e.g., crude tissue preparations, whole cells, preferably living cells, lysed cells, or cell fractions.

[0156] Included herein are antibodies which bind only a native P2X4 polypeptide, only denatured or otherwise non-native P2X4 polypeptide, or which bind both, as well as those having linear or conformational epitopes. Conformational epitopes sometimes can be identified by selecting antibodies that bind to native but not denatured P2X4 polypeptide.

[0157] Screening Assays

[0158] As used herein, the term "system" refers to a cell free in vitro environment and a cell-based environment such as a collection of cells, a tissue, an organ, or an organism. A system is contacted or introduced to a test molecule in a variety of manners, including adding molecules in solution and allowing them to interact with one another by diffusion, cell injection, and any administration routes in an animal. As used herein, the term "interaction" refers to an effect of a test molecule on a P2X4 nucleic acid, polypeptide, or variant thereof (collectively referred to as a "P2X4 molecule"), where the effect is sometimes binding between the test molecule and the nucleic acid or polypeptide, and often is an observable change in cells, tissue, or organism. There are many standard methods for detecting the presence or absence of interaction between a test molecule and a P2X4 nucleic acid or polypeptide, as described hereafter.

[0159] An interaction can be determined by labeling the test molecule and/or the P2X4 molecule, where the label is covalently or non-covalently attached to the test molecule or P2X4 molecule. The label is sometimes a radioactive molecule such as ^{125}I, ^{35}S, ^{14}C, or ^{3}H, which can be detected by direct counting of radioemmission or by scintillation counting. Also, enzymatic labels such as horseradish peroxidase, alkaline phosphatase, or luciferase may be utilized where the enzymatic label can be detected by determining conversion of an appropriate substrate to product. Also, presence or absence of an interaction can be determined without labeling. For example, a microphysiometer (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable

potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indication of an interaction between a test molecule and P2X4 (McConnell, H. M. et al., *Science* 257: 1906-1912 (1992)).

[0160] In cell-based systems, cells typically include a P2X4 nucleic acid or polypeptide or variants thereof and often are of mammalian origin, although the cell can be of any origin. Whole cells, cell homogenates, and cell fractions (e.g., cell membrane fractions) can be subjected to analysis. Where interactions between a test molecule with a P2X4 polypeptide or variant thereof are monitored, soluble and/or membrane bound forms of the polypeptide or variant may be utilized. Where membrane-bound forms of the polypeptide are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPS O), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

[0161] An interaction between two molecules can also be detected by monitoring fluorescence energy transfer (FET) (see, for example, Lakowicz et al., U.S. Pat. No. 5,631,169; Stavrianopoulos, et al., U.S. Pat. No. 4,868,103). A fluorophore label on a first, "donor" molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, "acceptor" molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the "donor" polypeptide molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the "acceptor" molecule label may be differentiated from that of the "donor". Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the "acceptor" molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorimeter).

[0162] In another embodiment, determining the presence or absence of an interaction between a test molecule and a P2X4 molecule can be effected by using real-time Biomolecular Interaction Analysis (BIA) (see, e.g., Sjolander, S. and Urbaniczky, C., *Anal. Chem.* 63: 2338-2345 (1991) and Szabo et al., *Curr. Opin. Struct. Biol.* 5: 699-705 (1995)). "Surface plasmon resonance" or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

[0163] In another embodiment, the P2X4 molecule or test molecules are anchored to a solid phase. The P2X4 molecule/test molecule complexes anchored to the solid phase can be detected at the end of the reaction. The target P2X4 molecule often is anchored to a solid surface, and the test molecule, which is not anchored, can be labeled, either directly or indirectly, with detectable labels discussed herein.

[0164] It may be desirable to immobilize a P2X4 molecule, an anti-P2X4 antibody, or test molecules to facilitate separation of complexed from uncomplexed forms of P2X4 molecules and test molecules, as well as to accommodate automation of the assay. Binding of a test molecule to a P2X4 molecule can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion polypeptide can be provided which adds a domain that allows a P2X4 molecule to be bound to a matrix. For example, glutathione-S-transferase/P2X4 fusion polypeptides or glutathione-S-transferase/target fusion polypeptides can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, Mo.) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target polypeptide or P2X4 polypeptide, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of P2X4 binding or activity determined using standard techniques.

[0165] Other techniques for immobilizing a P2X4 molecule on matrices include using biotin and streptavidin. For example, biotinylated P2X4 polypeptide or target molecules can be prepared from biotin-NHS(N-hydroxy-succinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

[0166] In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody).

[0167] In one embodiment, this assay is performed utilizing antibodies reactive with P2X4 polypeptide or test molecules but which do not interfere with binding of the P2X4 polypeptide to its test molecule. Such antibodies can be derivatized to the wells of the plate, and unbound target or P2X4 polypeptide trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes,

include immunodetection of complexes using antibodies reactive with the P2X4 polypeptide or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the P2X4 polypeptide or test molecule.

[0168] Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (see, for example, Rivas, G., and Minton, A. P., *Trends Biochem Sci Aug;*18(8): 284-7 (1993)); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (see, e.g., Ausubel, F. et al., eds. *Current Protocols in Molecular Biology,* J. Wiley: New York (1999)); and immunoprecipitation (see, for example, Ausubel, F. et al., eds. *Current Protocols in Molecular Biology,* J. Wiley: New York (1999)). Such resins and chromatographic techniques are known to one skilled in the art (see, e.g., Heegaard, N. H., *J. Mol. Recognit. Winter;* 11(1-6): 141-8 (1998); Hage, D. S., and Tweed, S. A., *J. Chromatogr. B Biomed. Sci. Appl. October* 10; 699 (1-2): 499-525 (1997)). Further, fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

[0169] In another embodiment, modulators of P2X4 expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of P2X4 mRNA or polypeptide evaluated relative to the level of expression of P2X4 mRNA or polypeptide in the absence of the candidate compound. When expression of P2X4 mRNA or polypeptide is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of P2X4 mRNA or polypeptide expression. Alternatively, when expression of P2X4 mRNA or polypeptide is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of P2X4 mRNA or polypeptide expression. The level of P2X4 mRNA or polypeptide expression can be determined by methods described herein for detecting P2X4 mRNA or polypeptide.

[0170] P2X4 Binding Partners

[0171] In another embodiment, binding partners that interact with a P2X4 molecule are detected. The P2X4 molecules can interact with one or more cellular or extracellular macromolecules, such as polypeptides, in vivo, and these molecules that interact with P2X4 molecules are referred to herein as "binding partners." Molecules that disrupt such interactions can be useful in regulating the activity of the target gene product. Such molecules can include, but are not limited to molecules such as antibodies, peptides, and small molecules. The preferred target genes/products for use in this embodiment are the P2X4 genes herein identified. In an alternative embodiment, the invention provides methods for determining the ability of the test compound to modulate the activity of a P2X4 polypeptide through modulation of the activity of a downstream effector of a P2X4 target molecule. For example, the activity of the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined, as previously described.

[0172] To identify compounds that interfere with the interaction between the target gene product and its cellular or extracellular binding partner(s), e.g., a substrate, a reaction mixture containing the target gene product and the binding partner is prepared, under conditions and for a time sufficient, to allow the two products to form complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner. Additionally, complex formation within reaction mixtures containing the test compound and normal target gene product can also be compared to complex formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

[0173] These assays can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the target gene product or the binding partner onto a solid phase, and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the target gene products and the binding partners, e.g., by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test compounds that disrupt preformed complexes, e.g., compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are briefly described below.

[0174] In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner, is anchored onto a solid surface (e.g., a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific for the species to be anchored can be used to anchor the species to the solid surface.

[0175] In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (e.g., by washing) and any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a

labeled antibody specific for the initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

[0176] Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

[0177] In an alternate embodiment of the invention, a homogeneous assay can be used. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared in that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, e.g., U.S. Pat. No. 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified.

[0178] Also, binding partners of P2X4 molecules can be identified in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Pat. No. 5,283,317; Zervos et al., *Cell* 72:223-232 (1993); Madura et al., *J. Biol. Chem.* 268: 12046-12054 (1993); Bartel et al., *Biotechniques* 14: 920-924 (1993); Iwabuchi et al., *Oncogene* 8: 1693-1696 (1993); and Brent WO94/10300), to identify other polypeptides, which bind to or interact with P2X4 ("P2X4-binding polypeptides" or "P2X4-bp") and are involved in P2X4 activity. Such P2X4-bps can be activators or inhibitors of signals by the P2X4 polypeptides or P2X4 targets as, for example, downstream elements of a P2X4-mediated signaling pathway.

[0179] A two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a P2X4 polypeptide is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified polypeptide ("prey" or "lsample") is fused to a gene that codes for the activation domain of the known transcription factor. (Alternatively the: P2X4 polypeptide can be the fused to the activator domain.) If the "bait" and the "prey" polypeptides are able to interact, in vivo, forming a P2X4-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the polypeptide which interacts with the P2X4 polypeptide.

[0180] Identification of Candidate Therapeutics

[0181] Candidate therapeutics for reducing fat deposition or treating NIDDM are identified from a group of test molecules that interact with a P2X4 nucleic acid or polypeptide. Test molecules often are ranked according to the degree with which they interact or modulate (e.g., agonize or antagonize) DNA replication and/or processing, RNA transcription and/or processing, polypeptide production and/or processing, and/or function of P2X4 molecules, for example, and then top ranking modulators are selected. Also, pharmacogenomic information described herein can determine the rank of a modulator. Candidate therapeutics typically are formulated for administration to a subject.

[0182] Therapeutic Treatments

[0183] Formulations or pharmaceutical compositions typically include in combination with a pharmaceutically acceptable carrier a compound, an antisense nucleic acid, an siRNA molecule capable of inhibiting the expression of P2X4 or, optionally, any of its transcripts, a ribozyme, an antibody, or a binding partner that interacts with a P2X4 polypeptide, a P2X4 nucleic acid, or a fragment thereof. The formulated molecule that is formulated may be one that is identified by a screening method described above. Also, formulations may comprise a P2X4 polypeptide, P2X4 nucleic acid, or fragment thereof and a pharmaceutically acceptable carrier. As used herein, the term "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

[0184] A pharmaceutical composition is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerin, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[0185] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile

and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[0186] Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0187] Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, e.g., gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[0188] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

[0189] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art. Molecules can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[0190] In one embodiment, active molecules are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. Materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Pat. No. 4,522,811.

[0191] It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

[0192] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_5O/ED_{50}. Molecules which exhibit high therapeutic indices are preferred. While molecules that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[0193] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such molecules lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any molecules used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

[0194] As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage)

ranges from about 0.001 to 30 mg/kg body weight, sometimes about 0.01 to 25 mg/kg body weight, often about 0.1 to 20 mg/kg body weight, and more often about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, sometimes between 2 to 8 weeks, often between about 3 to 7 weeks, and more often for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors may influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments.

[0195] With regard to polypeptide formulations, featured herein is a method for reducing fat deposition, alleviating obesity and/or alleviating NIDDM in a subject, which comprises contacting a P2X4 protein with one or more cells of a subject in need thereof, wherein the P2X4 protein is encoded by a P2X4 nucleotide sequence which comprises a polynucleotide sequence selected from the group consisting of (a) the polynucleotide sequence of SEQ ID NO: 1; (b) a polynucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; (c) a polynucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3 or a polynucleotide sequence 90% identical to the nucleotide sequence of SEQ ID NO:1; and (d) a fragment of one of the foregoing polynucleotide sequences, where contacting the one or more cells of the subject with the P2X4 protein reduces fat deposition, alleviates obesity and/or alleviates NIDDM. The P2X4 protein often is administered to a subject prognosed as being at risk of fat deposition, obesity and/or NIDDM or is diagnosed as having obesity or NIDDM before the protein is administered in vivo (e.g., injected into the subject), ex vivo (e.g., cells from the subject are contacted with the protein in a petri dish and the contacted cells then are returned to the subject), or in vitro (e.g., cells from the subject are contacted with the protein in a petri dish to observe the effect of the protein on the cells). The subject often is a human.

[0196] For antibodies, a dosage of 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg) often is utilized. If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg often is appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration often is possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank et al., *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193 (1997).

[0197] Antibody conjugates can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a polypeptide such as tumor necrosis factor, .alpha.-interferon, .beta.-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1 "), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors. Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Pat. No. 4,676,980.

[0198] For compounds, exemplary doses include milligram or microgram amounts of the compound per kilogram of subject or sample weight, for example, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated. Candidate therapeutic molecules identified in screening methods described previously sometimes are administered to cells of a subject in vivo, ex vivo, or in vitro to reduce fat deposition, alleviate obesity and/or alleviate NIDDM in the subject, and the subject often is prognosed or diagnosed as having such conditions before the candidate therapeutic is contacted with the cells.

[0199] P2X4 nucleic acid molecules can be inserted into vectors and used in gene therapy methods for reducing fat deposition or treating NIDDM. Featured herein is a method for reducing fat deposition, alleviating obesity and/or alleviating NIDDM in a subject, which comprises contacting a P2X4 nucleic acid with one or more cells of a subject in need thereof, wherein the P2X4 nucleic acid comprises a polynucleotide sequence selected from the group consisting of (a) the polynucleotide sequence of SEQ ID NO:1; (b) a polynucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; (c) a polynucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3 or a polynucleotide sequence 90% identical to the nucleotide sequence of SEQ ID NO:1; and (d) a fragment of one of the foregoing polynucleotide sequences, where contacting the one or more cells of the subject with the P2X4 protein reduces fat deposition, alleviates obesity and/or alleviates NIDDM. The P2X4 nucleic acid often is administered to a subject prognosed as being at risk of fat deposition, obesity and/or NIDDM or is diagnosed as having obesity or NIDDM before

the nucleic acid is administered in vivo, ex vivo, or in vitro. The subject often is a human.

[0200] Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Pat. No. 5,328,470) or by stereotactic injection (see e.g., Chen et al., (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). Pharmaceutical preparations of gene therapy vectors can include a gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells (e.g., retroviral vectors) the pharmaceutical preparation can include one or more cells which produce the gene delivery system. Examples of gene delivery vectors are described herein.

[0201] Pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration. Pharmaceutical compositions of active ingredients can be administered by any of the paths described herein for therapeutic and prophylactic methods for reducing fat deposition or treating NIDDM. With regard to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from pharmacogenomic analyses described herein. As used herein, the term "treatment" is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

[0202] A prophylactic agent often is administered prior to the manifestation of symptoms characteristic of the P2X4 aberrance, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of P2X4 aberrance, for example, a P2X4 molecule, P2X4 agonist, or P2X4 antagonist agent can be used for treating the subject. The appropriate agent often is determined based on screening assays described herein.

[0203] As discussed, successful treatment of P2X4 disorders can be brought about by techniques that serve to inhibit the expression or activity of target gene products. For example, compounds (e.g., an agent identified using an assays described above) that exhibit negative modulatory activity can be used in accordance with the invention to prevent and/or ameliorate fat deposition or NIDDM. Such molecules can include but are not limited to peptides, phosphopeptides, small organic or inorganic molecules, siRNA molecules or antibodies (including, for example, polyclonal, monoclonal, humanized, anti-idiotypic, chimeric or single chain antibodies, and Fab, $F(ab')_2$ and Fab expression library fragments, scFV molecules, and epitope-binding fragments thereof).

[0204] Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used in accordance with the invention to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above.

[0205] It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular polypeptide, it can be preferable to co-administer normal target gene polypeptide into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

[0206] Another method by which nucleic acid molecules may be utilized in treating or preventing a disease characterized by P2X4 expression is through the use of aptamer molecules specific for P2X4 polypeptide. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically bind to polypeptide ligands (see, e.g., Osborne, et al., *Curr. Opin. Chem. Biol.* 1(1): 5-9 (1997); and Patel, D. J., *Curr. Opin. Chem. Biol. Jun;* 1(1): 32-46 (1997)). Since nucleic acid molecules may in many cases be more conveniently introduced into target cells than therapeutic polypeptide molecules may be, aptamers offer a method by which P2X4 polypeptide activity may be specifically decreased without the introduction of drugs or other molecules which may have pluripotent effects.

[0207] Antibodies can be generated that are both specific for target gene product and that reduce target gene product activity. Such antibodies may, therefore, by administered in instances whereby negative modulatory techniques are appropriate for the treatment of P2X4 disorders. For a description of antibodies, see the Antibody section above.

[0208] In circumstances where injection of an animal or a human subject with a P2X4 polypeptide or epitope for stimulating antibody production is harmful to the subject, it is possible to generate an immune response against P2X4 through the use of anti-idiotypic antibodies (see, for example, Herlyn, D., *Ann. Med.*;31(1): 66-78 (1999); and Bhattacharya-Chatterjee, M., and Foon, K. A., *Cancer Treat. Res.;* 94: 51-68 (1998)). If an anti-idiotypic antibody is introduced into a mammal or human subject, it should stimulate the production of anti-anti-idiotypic antibodies, which should be specific to the P2X4 polypeptide. Vaccines directed to a disease characterized by P2X4 expression may also be generated in this fashion.

[0209] In instances where the target antigen is intracellular and whole antibodies are used, internalizing antibodies may be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be

administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (see e.g., Marasco et al., *Proc. Natl. Acad. Sci. USA* 90: 7889-7893 (1993)).

[0210] P2X4 molecules and compounds that inhibit target gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to prevent, treat or ameliorate P2X4 disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disorders.

[0211] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}. Compounds that exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[0212] Data obtained from cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[0213] Another example of effective dose determination for an individual is the ability to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays may utilize antibody mimics and/or "biosensors" that have been created through molecular imprinting techniques. The compound which is able to modulate P2X4 activity is used as a template, or "imprinting molecule", to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated "negative image" of the compound and is able to selectively rebind the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell, R. J. et al., *Current Opinion in Biotechnology* 7: 89-94 (1996) and in Shea, K. J., *Trends in Polymer Science* 2: 166-173 (1994). Such "imprinted" affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis, G. et al., *Nature* 361. 645-647 (1993). Through the use of isotope-labeling, the "free" concentration of compound which modulates the expression or activity of P2X4 can be readily monitored and used in calculations of IC_{50}. Such "imprinted" affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes can be readily assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC_{50}. A rudimentary example of such a "biosensor" is discussed in Kriz, D. et al., *Analytical Chemistry* 67: 2142-2144 (1995).

[0214] Provided herein are methods of modulating P2X4 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a P2X4 or agent that modulates one or more of the activities of P2X4 polypeptide activity associated with the cell. An agent that modulates P2X4 polypeptide activity can be an agent as described herein, such as a nucleic acid or a polypeptide, a naturally-occurring target molecule of a P2X4 polypeptide (e.g., a P2X4 substrate or receptor), a P2X4 antibody, a P2X4 agonist or antagonist, a peptidomimetic of a P2X4 agonist or antagonist, or other small molecule.

[0215] In one embodiment, the agent stimulates one or more P2X4 activities. Examples of such stimulatory agents include active P2X4 polypeptide and a nucleic acid molecule encoding P2X4. In another embodiment, the agent inhibits one or more P2X4 activities. Examples of such inhibitory agents include antisense P2X4 nucleic acid molecules, anti-P2X4 antibodies, and P2X4 inhibitors. These modulatory methods can be performed in vitro (e.g., by culturing the cell with the agent) or, alternatively, in vivo (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a P2X4 polypeptide or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) P2X4 expression or activity. In another embodiment, the method involves administering a P2X4 polypeptide or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted P2X4 expression or activity.

[0216] Stimulation of P2X4 activity is desirable in situations in which P2X4 is abnormally downregulated and/or in which increased P2X4 activity is likely to have a beneficial effect. For example, stimulation of P2X4 activity is desirable in situations in which a P2X4 is downregulated and/or in which increased P2X4 activity is likely to have a beneficial effect. Likewise, inhibition of P2X4 activity is desirable in situations in which P2X4 is abnormally upregulated and/or in which decreased P2X4 activity is likely to have a beneficial effect.

[0217] With regard to treatment of NIDDM, featured herein are methods of causing or inducing a desired biological response in an individual comprising the steps of: providing or administering to an individual a composition comprising a P2X4 nucleic acid, P2X4 protein, anti-P2X4 antibody, P2X4 interacting molecule or another therapeutic composition described herein, where the biological response

is selected from the group consisting of: (a) modulating circulating (e.g., blood, serum or plasma) levels (e.g., concentration) of glucose (e.g., lowering glucose levels); (b) increasing cell or tissue sensitivity to insulin, particularly muscle, adipose, liver or brain; (c) inhibiting the progression from impaired glucose tolerance to insulin resistance; (d) increasing glucose uptake in skeletal muscle cells; (e) increasing glucose uptake in adipose cells; (f) increasing glucose uptake in neuronal cells; (g) increasing glucose uptake in red blood cells; (h) increasing glucose uptake in the brain; and (i) significantly reducing the postprandial increase in plasma glucose following a meal, particularly a high carbohydrate meal (e.g., about 2-fold, about 10-fold and about 50-fold reductions).

[0218] In further embodiments, a pharmaceutical or physiologically acceptable composition is used as an insulin sensitiser. In other embodiments, a pharmaceutical or physiologically acceptable composition is used in a method to improve insulin sensitivity in subjects having NIDDM in combination with insulin therapy, or in an alternative embodiment, without insulin therapy. In other embodiments, a pharmaceutical or physiologically acceptable composition described herein is utilized in a method of treating individuals with gestational diabetes. "Gestational diabetes" refers to the development of diabetes in an individual during pregnancy, usually during the second or third trimester of pregnancy. In further embodiments, a pharmaceutical or physiologically acceptable composition described herein is used in a method for treating individuals with impaired fasting glucose (IFG). "Impaired fasting glucose" (IFG) is a condition in which fasting plasma glucose levels in an individual are elevated but not diagnostic of overt diabetes, i.e. plasma glucose levels of less than 126 mg/dl and greater than or equal to 110 mg/dl. In other embodiments, a pharmaceutical or physiologically acceptable composition described herein is used in a method for treating and preventing impaired glucose tolerance (IGT) in an individual. By providing therapeutics and methods for reducing or preventing IGT (i.e., for normalizing insulin resistance), the progression to NIDDM can be delayed or prevented. Furthermore, by providing therapeutics and methods for reducing or preventing insulin resistance, provided are methods for reducing and/or preventing the appearance of insulin-resistance syndrome (IRS).

[0219] In further embodiments, a pharmaceutical or physiologically acceptable composition described herein is used in a method for treating a subject having polycystic ovary syndrome (PCOS). PCOS is among the most common disorders of pre-menopausal women, affecting 5-10% of this population. Insulin-sensitizing agents (e.g., troglitazone) have been shown to be effective in PCOS and improve defects in insulin action, insulin secretion, ovarian steroidogenosis and fibrinolysis (Ehrman et al. (1997) J Clin Invest 100:1230) in insulin-resistant humans. Accordingly, provided are methods for reducing insulin resistance, normalizing blood glucose thus treating and/or preventing PCOS.

[0220] In other embodiments, a pharmaceutical or physiologically acceptable composition described herein is used in a method for treating a subject having insulin resistance. In further embodiments, a subject having insulin resistance is treated according to the methods described herein to reduce, alleviate, or cure the insulin resistance. As insulin resistance often is associated with infections and cancer, reducing, alleviating and/or preventing insulin resistance according to the methods of described herein can prevent or reduce infections and cancer. In another embodiment, methods described herein are used to prevent the development of insulin resistance in a subject (e.g., those known to have an increased risk of developing insulin resistance).

[0221] Any of the above-described tests or other tests known in the art can be used to determine that a subject is insulin resistant, and a patient thus diagnosed then is treated according to the methods described herein to reduce or cure insulin resistance. Alternatively, methods described herein also can be used to prevent the development of insulin resistance in a subject, e.g., those known to have an increased risk of developing insulin-resistance.

[0222] The examples set forth below are intended to illustrate but not limit the invention.

EXAMPLES

[0223] In the following studies a group of subjects were selected according to specific parameters relating to fat deposition. Nucleic acid samples obtained from individuals in the study group were subjected to genetic analysis, which identified associations between central obesity and certain polymorphic regions in the P2X4 gene on chromosome 12. Polymorphic variations identified as being associated with central obesity were further screened in subjects with NIDDM to determine if they are also associated with the development of diabetes. Methods are described for producing P2X4 polypeptide and P2X4 polypeptide variants in vitro or in vivo, P2X4 nucleic acids or polypeptides and variants thereof are utilized for screening test molecules for those that interact with P2X4 molecules. Test molecules identified as interactors with P2X4 molecules and P2X4 variants are further screened in vivo to determine whether they can reduce fat deposition or treat NIDDM.

Example 1

Sample Selection

[0224] In addition to simple clinical measurements, dual x-ray absorbtiometry (DEXA) was utilized to determine fat content in subjects for the genetic analysis. Central fat was the primary target variable, and data were collected using a Hologic QDR 4500 DEXA system. The central region for central fat determinations was defined as the region extending from the superior surface of the second lumbar vertebra extending inferiorly to the inferior surface of the fourth lumbar vertebra and laterally to the inner aspect of the ribcage. The amount of central fat and percent central fat was automatically calculated by the equipment and downloaded into a database.

[0225] Waist and hip measurements were generated while subjects were wearing underclothes and standing with their arms by their sides. A tape measure was utilized for these measurements, and care was taken to ensure that the tape was resting on the skin and not tight. Waist circumference was measured to the nearest centimeter at the narrowest point between the iliac crest and the lower edge of the ribs. Hip circumference was measured to the nearest centimeter at the widest point below the iliac crest.

[0226] Sample selection was restricted to female twins followed by the St. Thomas Hospital in England and the Royal North Shore Hospital in Australia. It was estimated that 552 dizygotic sibling pairs would yield statistical results of reasonable power. A further 272 unrelated individuals selected from monozygotic twin pairs were added to the sample set to increase the probability for detecting associations and also for testing gene-environment interactions. The study group was selected from this combination of dizygotic and monozygotic sibling pairs, referred to as the "selection group."

[0227] Central fat measurements and triglyceride measurements were chosen as primary target phenotypes and twin pairs were selected from the selection group for extreme discordance and concordance. Specifically, DEXA measurements and triglyceride measurements (colorimetric enzymatic method: glycerol-3-phosphate-oxidase, peroxidase, PAP (Roche), CV%=2.6, reference range less than 2.5 mmol/L) for each individual in the selection group were arranged in ascending order, and individuals in the top and lower tenth percentile were chosen from each distribution. A small subset of individuals falling in the middle range of each distribution was chosen as a control group. In addition to primary phenotype trait information, samples for inclusion in the study group were selected based on data coverage for the following secondary phenotypes recorded by each individual: BMI, insulin resistance, high density lipoprotein in serum, waist, lipoprotein(a) in serum, insulin, hip, and waist/hip ratio.

[0228] Also, presence of diabetes, thyroid disease, and renal disease reported by each individual were primary criteria for excluding subjects from the study group. Also, insulin levels greater than 7.1 μU/ml (Microparticle Enzyme Immunoassay from Abbott Laboratories Diagnostics Division (μU/ml=pmol/L×7.175) and creatine levels greater than 160 mmol/L (measured by Jaffe method: calorimetric test in which creatine reacts with picric acid in an alkaline solution to form a yellow-red colored complex) were also used as exclusion criteria as they are indicative of these diseases. Further excluded were pairs discordant for menopausal status, twin pairs where one or both of the twins were taking lipid lowering medication, non-fasting subjects (less than eight hours eating), and twin pairs including subjects treated with beta-blockers, thiazide diuretics, or exogeneous estrogen.

[0229] Selecting among dizygotic and monozygotic twins for extreme discordance or concordance for the primary phenotypes minimized complications associated with bivariate ranking. After applying exclusion criteria, 253 monozygotic subjects were available for inclusion, which fell short of the target population of 276. In reaction to this situation, the extreme 201 subjects were selected from the 253 subjects and the desired numbers were reached by adding monozygotic unrelated individuals with data for central fat only, and unrelated individuals from the dizygotic cohort with data for triglycerides only. Samples available for final selection for the 552 dizygotic pairs included 178 pairs extreme for both traits, 205 extreme for triglycerides only, and 208 for central fat only.

[0230] In this test population, coverage for the secondary phenotypes ranged from 67% to 90%. In total, 61% of subjects had coverage for all primary and secondary phenotypes. A broad age spectrum was also represented, and numbers of pre-menopausal and post-menopausal subjects were relatively evenly distributed.

Example 2

Association of Polymorphic Variations to Fat Deposition

[0231] Blood samples were taken from individuals in the study population described in Example 1. Genomic DNA was extracted from these blood samples using standard techniques (BACC2 DNA extraction kit (Nucleon Biosciences)) and subjected to analysis. Based upon a background linkage study and fine mapping analysis by microsatellite markers, it was postulated that genetic elements linked to central fat deposition were located on the 12q24 region of chromosome twelve. One of the genes located in this region encoded P2X4.

[0232] Whole genome linkage scans were performed for the purpose of identifying genomic regions likely to harbor genes with a major contribution to deposition of central fat. The linkage scans were performed using highly polymorphic microsatellite markers (Reed et al., *Nat Genet* 7:390-5 (1994)) and DNA samples obtained from 1100 Caucasian female twin pairs from the UK. Samples selected for inclusion in the study cohort encompassed a broad spectrum of phenotypic trait values, ranging from lean to obese subjects. Initial studies were carried out using 400 commercially available microsatellite markers derived from the Genethon linkage map, with an average genomic spacing between markers of approximately 10 cM (ABI Prism linkage mapping set, version 2 from PE Applied Biosystems).

[0233] Multipoint nonparametric linkage analysis was performed using MAPMAKER/SIBS (Kruglyak & Lander, *Amer. J. Human Genetics* 57:439-454 (1995)). A bioinformatics infrastructure and software packages described in WO 00/51053 were used in the linkage study to record marker positions, store data and generate data files. Output from these systems was then used with relevant application software to perform the statistical analysis.

[0234] Genotyping reactions were generally carried out in microtitre plates (384-well, reaction volume 5 μl), containing 12.5 ng of DNA from study subjects was amplified using PCR and sequence specific oligonucleotide primers labeled with 6-FAM™, HEX™, or NED™ fluorescent dyes. PCR products were analyzed by electrophoresis in a polyacrylamide denaturing gel, with an ABI PRISM™ GENESCAN® 400HD ROX labeled size standard in each lane on an ABI model 377 analyzer (Applied Biosystems, Foster City, Calif.). For genotyping, the chosen markers were divided into two groups (panels) so that the analysis of all of the markers could be performed in two electrophoresis runs of each sample. Consequently, there was no overlap of fragment sizes in any one dye for either of the panels. Genotype analysis was performed using ABI PRISM™ GENESCAN® software (version 3.0), and genotyped manually using ABI PRISM™ Genotyper 2.0. Results were input into a database and binned by marker. The results were quality checked, ensuring consistent inheritance within families. Families that were found to have consistent pedigree problems were excluded from the analysis set.

[0235] The ordering of genetic mapping markers (i.e. microsatellite markers) was relatively stable in the region analyzed according to the Unified Data Base for Human Genome Mapping, Weizmann Institute of Science (UDB) and National Center for Biotechnology Information, National Institutes of Health (NCBI) assemblies during the duration of the study. Conversion of genetic to physical positions for strategic microsatellite markers was performed using UDB and NCBI as the reference standards. Comparisons of the identity and positioning of genomic contigs in the region also were made between UDB and NCBI and provided relatively good agreement. A comparison of the positioning of all identified and predicted genes within the region also was made between NCBI (build 22) and Joint Project between European Bioinformatics Institute and the Sanger Centre (ENSEMBL).

[0236] Microsatellite marker analysis showed linkage on the long arm of chromosome 12 to central fat deposition, percent central fat and total fat in the region spanning 125 cM to 155 cM, with a peak non parametric Z score of 3.6 for central fat. The region was narrowed further to identify the chromosomal interval 12q24 as being the primary region harboring genes contributing to central fat deposition using the following highly polymorphic microsatellite markers: D12S86, D12S1612, D12S1614, D12S340, D12S324, D12S1675, D12S1679, D12S1659 and D12S97.

[0237] The chromosome 12q24 region then was analyzed using single nucleotide polymorphisms to identify genes in the region that regulate central fat deposition. Potential polymorphisms in the P2X4 polynucleotide were identified in a publicly available SNP database (see http address www.ncbi.nlm.nih.gov/SNP) and were verified in a group other than the study group. Polymorphisms verified as statistically significant SNPs (i.e., minor allele represented in more than 10% of the population) were genotyped in the study population to determine associations with fat deposition. A procedure for detecting polymorphisms was utilized in the verification and genotyping studies, described hereafter. Table 1 shows the majority of polymorphisms subjected to genotype analysis, where "np" indicates that the SNP is not present in the sequence of SEQ ID NO:1 and "-" indicates that there is no reference SNP ID number. Reference SNP number rs1044251 is now called rs25643, and corresponds to the same SNP described previously by rs1044251 (see http address www.ncbi.nlm.nih.gov/SNP).

TABLE 1

Reference SNP ID	Position in SEQ ID NO: 1	Position in Ensembl Database	Position in NCBI Database	Allelic Variability
rs930187	7337	4577	7336	A/G
rs1182945	11030	8270	11029	T/A
rs1044251 (rs25643)	15847	13087	15846	T/C
rs725135	17338	14577	17336	T/C
rs25644	21708	19093	21699	A/G
rs1151880	22713	20098	22704	C/T
—	14744	11984	11744	C/T

[0238] Assays for Verifying and Genotyping SNPs

[0239] An assay utilized for determining whether a polymorphic variation was present in a nucleic acid sample involved a sequencing by synthesis procedure. DNA polymerase, ATP sulfurylase, luciferase, apyrase, luciferin, and adenosine 5'-phosphosulfate (APS) were required, and in the process, one dNTP was added to an extension oligonucleotide at a time and then degraded if not incoroporated in the synthesized strand. Incorporation of a dNTP to the end of the extension oligonucleotide was detected by light emission.

[0240] The assay was carried out by first amplifying a region of interest in the sample by using a polymerase chain reaction (PCR) that incorporated the primers set forth in Table 2.

TABLE 2

Reference SNP ID	Position in SEQ ID NO:1	Forward PCR primer	SEQ ID NO:	Reverse PCR primer	SEQ ID NO:
rs930187	7337	AACCCTCCCTCAAAAAAGAC		GGTTTCATGCCACTCATTATCA	
rs1182945	11030	GGAAAACAAGAACTCCAGAAACT		CCCTTCCTCCTGACTTTTATG	
rs1044251 (rs25643)	15847	TGCCACCCTTGTGCTTGTA		AAAACAAAACAAAAGGCACTTC	
rs725135	17338	CTGAGGCAGTCGGATCATC		CCAAGCTAGAGTGCAGAAGAG	
rs25644	21708	GAGGTTTTCGTCGCTCTGAT		AGGGCTGGGTCAGGAGAG	
rs1151880	22713	TCACTCCAGTCTCTGCCTCTC		GCCCTAAATTCAGTGACTGATAT	

[0241] A typical PCR reaction included 14.24 μl of water, 2.23 μl of PCR buffer, 1.38 μl of 1.5 mM $MgCl_2$, 1.12 μl of 0.125 mM dNTPs, 0.45 μl of the forward primer at a 0.2 μM concentration, 0.45 μl of the reverse primer at a 0.2 μM concentration, 0.13 μl of Taq polymerase (0.003 U/μl), and 2.3 μl of DNA sample at a 0.2 ng/μl concentration, for a total volume of 22.3 μl. The PCR reaction normally was carried out using one step at 95° C. for 10 minutes; 45 cycles at 95° C. for 30 seconds, 60° C. for 45 seconds, and 72° C. for 45 seconds; one step at 72° C. for 5 minutes; and then finalizing the reaction at 22° C.

[0242] After the PCR reaction was completed, an extension oligonucleotide was hybridized to the PCR product. Extension oligonucleotides are reported in Table 3.

TABLE 3

Position in SEQ ID NO:1	Extension Oligonucleotide	SEQ ID NO:
7337	AAAAAAAGACCGAACAC	
11030	CATTAGCTCGGTCTCC	
15847	GCTTTCAACGGGTC	
17338	TAGCTGGGATTACAGAC	
21708	CATGTCCTGGAAAC	
22713	GCCTTCTGCCTTGTA	

[0243] The extension oligonucleotide was complementary to the amplified target up to but not including the polymorphism, and was extended enzymatically one or a few bases through the polymorphic site. In the extension phase of the assay, a single dNTP was added to the reaction, and pyrophosphate was generated if the dNTP was added to the extension oligonucleotide. ATP sulfurylase present in the reaction mixture utilized the pyrophosphate in conjunction with APS to generate ATP. ATP drove the luciferase-catalyzed conversion of luciferin to oxyluciferin, which generated the release of visible light measured by a CCD camera. A graphic representation was generated showing a peak corresponding to the amount of light emitted, where the light was proportional to the amount of nucleotide incorporated into the extension oligonucleotide. dATP was not added to the reaction, and instead, was replaced by dATPyS, which was not turned over by luciferase. Apyrase was added to the reaction to degrade unicorporated dNTP and ATP sulfurylase-generated ATP, and when the apyrase reaction was complete, another dNTP was optionally added to the reaction for another extension phase.

[0244] An alternative assay involved a MassARRAY™ system (Sequenom, Inc.), which was utilized to perform SNP genotyping in a high-throughput fashion. This genotyping platform was complemented by a homogeneous, single-tube assay method (hME™ or homogeneous MassEXTEND™ (Sequenom, Inc.)) in which two genotyping primers anneal to and amplify a genomic target surrounding a polymorphic site of interest. A third primer (the MassEXTEND™ primer), which is complementary to the amplified target up to but not including the polymorphism, then was enzymatically extended one or a few bases through the polymorphic site and then terminated.

[0245] For each polymorphism, SpectroDESIGNER™ software (Sequenom, Inc.) was used to generate a set of PCR primers and a MassEXTEND™ primer was used to genotype the polymorphism. Table 4 shows PCR primers and Table 5 shows extension primers used for analyzing polymorphisms. The initial PCR amplification reaction was performed in a 5 μl total volume containing 1×PCR buffer with 1.5 mM $MgCl_2$ (Qiagen), 50 μM each of dATP, dGTP, dCTP, dTTP (Gibco-BRL), 2.5 ng of genomic DNA, 0.1 units of HotStar DNA polymerase (Qiagen), and 200 nM each of forward and reverse PCR primers specific for the polymorphic region of interest.

TABLE 4

Reference SNP ID	Position in SEQ ID NO:1	Forward PCR primer	SEQ ID NO:	Reverse PCR primer	SEQ ID NO:
rs930187	7337	ACGTTGGATGACCCTCCCTCAAAAA AAGAC		ACGTTGGATGCATGCCACTCA TTATCATGTG	
rs1044251 (rs25643)	15847	ACGTTGGATGTTGTAGGAGTCTCAA CAGGC		ACGTTGGATGACGTGTGTGTC ATCCTCCAC	
rs25644	21708	ACGTTGGATGGCCAAAATAGTGGA GAACGC		ACGTTGGATGAGTTGACCTGG ATGCCCATG	
rs1151880	22713	ACGTTGGATGTCCTGTCATGAGGCC TTCTG		ACGTTGGATGGGCCCTAAATT CAGTGACTG	
rs208307	np	TCCTGGGAAAGAGACAGATC		GTGATGTTTAAACCTGGCAG	
rs1621388	np	TGCCTGGCTTCAGTAAGGAC		CGCTGGAGGATCCGGAAAGA	
rs2686384	np	TGGTTCCTGAGTTTAAAGCC		AAAAACCAGCAGCTTTCTAC	
rs1182945	11030	CTCAAGAGAAGGGTCTTGTG		CCTCCTGACTTTTATGGCTC	
rs1653586	np	TCCCGACCCTTTCTACTATG		TGTGGAAAAACACGAAAGGC	
rs1653589	np	TATATACAGGCGGGCATAGG		TTCACAGCTTCAGGCATGTG	
rs1063843	np	GAAACAGAAGTGACTGCCTG		CACTGCCCTTCCCACTGAG	

[0246] Samples were incubated at 95° C. for 15 minutes, followed by 45 cycles of 95° C. for 20 seconds, 56° C. for 30 seconds, and 72° C. for 1 minute, finishing with a 3 minute final extension at 72° C. Following amplification, shrimp alkaline phosphatase (SAP) (0.3 units in a 2 μl

volume) (Amersham Pharmacia) was added to each reaction (total reaction volume was 7 μl) to remove any residual dNTPs that were not consumed in the PCR step. Samples were incubated for 20 minutes at 37° C., followed by 5 minutes at 85° C. to denature the SAP.

[0247] Once the SAP reaction was complete, a primer extension reaction was initiated by adding a polymorphism-specific MassEXTEND™ primer cocktail to each sample. Each MassEXTEND™ cocktail included a specific combination of dideoxynucleotides (ddNTPs) and deoxynucleotides (dNTPs) used to distinguish polymorphic alleles from one another. In Table 5, dNTPs are shown and the fourth nucleotide not shown is the ddNTP (e.g., in the first row A, C and T are dNTPs and G is the ddNTP).

TABLE 5

Reference SNP ID	Position in SEQ ID NO:1	Extend Probe	SEQ ID NO:	Termination Mix
rs930187	7337	CCTCAAAAAAAGACCGAACAC		A,C,T
rs1044251 (rs25643)	15847	GCGTAGCTTTCAACGGGTC		A,C,G
rs25644	21708	ACGGCCATGTCCTGGAAAC		A,C,G
rs1151880	22713	GAAGAAAAGACACAGAGACAC		A,C,T
rs208307	ns	TGGCAGGATGTTTCTCCTGT		A,C,T
rs1621388	ns	AGGATCCGGAAAGAGTTTCC		A,C
rs2686384	ns	CAGCTTTCTACTATATTCATGTAA		A,C
rs1182945	ns	GCTCATTAGCTCGGTCTCC		C,T
rs1653586	ns	GGAAAAACACGAAAGGCTTAGTT		C,T
rs1653589	ns	TCAGGCATGTCATCAGTTAGAGCC		A,C,T
rs1063843	ns	CCTTCCCACTGAGCACATC		A,C,T

[0248] The MassEXTEND™ reaction was performed in a total volume of 9 μl, with the addition of 1× ThermoSequenase buffer, 0.576 units of ThermoSequenase (Amersham Pharmacia), 600 nM MassEXTEND™ primer, 2 mM of ddATP and/or ddCTP and/or ddGTP and/or ddTTP, and 2 mM of dATP or dCTP or dGTP or dTTP. The dideoxy (dd) nucleotide used in the assay was complementary to the nucleotide at the polymorphic site in the amplicon. Samples were incubated at 94° C. for 2 minutes, followed by 45 cycles of 5 seconds at 94° C., 5 seconds at 52° C., and 5 seconds at 72° C.

[0249] Following incubation, samples were desalted by adding 16 μl of water (total reaction volume was 25 μl), 3 mg of SpectroCLEAN™ sample cleaning beads (Sequenom, Inc.) and allowed to incubate for 3 minutes with rotation. Samples were then robotically dispensed using a piezoelectric dispensing device (SpectroJET™ (Sequenom, Inc.)) onto either 96-spot or 384-spot silicon chips containing a matrix that crystallized each sample (SpectroCHIP™ (Sequenom, Inc.)). Subsequently, MALDI-TOF mass spectrometry (Biflex and Autoflex MALDI-TOF mass spectrometers (Bruker Daltonics) can be used) and SpectroTYPER RT™ software (Sequenom, Inc.) were used to analyze and interpret the SNP genotype for each sample.

[0250] SNP Verification

[0251] Polymorphisms identified in the publicly available database were verified by detecting the presence or absence of each polymorphism across six individuals from Sweden (including PCR negative control and one sequence primer extension control). Where a polymorphism was present in two or more of the individuals, the polymorphism was designated as a statistically significant SNP and genotyped across the test population. Where the polymorphism was not identified in any of the six individuals, it was further examined in a population of thirty Caucasian blood donors from Sweden. In this group of thirty individuals, a polymorphism having a frequency of 10% or greater was designated as a statistically significant SNP and genotyped across the test population. The probability of not identifying a minor allele variant represented in 10% or more of a population was calculated as being about 0.2% when samples from 30 individuals are analyzed, where it was estimated that 19% of individuals in the total population would be carriers for the minor allele assuming a large population and no selection pressure.

[0252] The following polymorphisms reported in the dbSNP database were identified as being polymorphic (i.e., statistically significant) in the verification studies: rs208307, rs1063843, rs1182945, rs1621388, rs1653586, rs1653589, rs2686384, rs1044251 (rs25643), rs25644, and rs1151880. Polymorphisms reported in the dbSNP database as rs96064, rs1044249, rs2303998, and rs2288690 were identified as not being polymorphic, where the first three were tested in seventeen individuals and where the fourth was tested in 30 individuals described in Example 3. Also, a SNP at position 14744 in SEQ ID NO:1, which was not reported in the dbSNP database, was identified in sequencing/genotyping studies described in Example 3 using the primers referred to in Table 8 as "Exons 3&4" primers.

[0253] Genotype Analysis

[0254] Among the verified SNPs, Table 6 depicts three SNPs that were strongly associated with fat deposition.

Allele frequency is noted in the second column and the allele indicated in bold type is the allele associated with increased central fat deposition. Statistical significance of each association was determined by the computer program QTDT (Abecasis et al., *Amer. J Human Genetics,* 66: 279-292 (2000)). Transmission/disequilibrium tests (TDT), population stratification tests, and total association tests were utilized in this computer program. Because population stratification could have lead to spurious results from total association testing, association testing was considered only when there was no evidence of population stratification.

TABLE 6

SNP Position in SEQ ID NO: 1	Allele Frequency	Statistical Significance
11030	**A = 0.67**	p = 0.0168
	T = 0.33	
15847	**T = 0.55**	p = 0.0317
	C = 0.45	
17338	**C = 0.76**	p = 0.0146
	T = 0.24	
14744	C = 0.55	p = 0.0000018
	T = 0.45	

[0255] SNP 14744 was originally genotyped over 30 dizygotic and monozygotic twins from the St. Thomas and Royal North Shore samples described in Example 1, and was found to be associated with fat deposition. Upon genotype analysis over the remaining St. Thomas and Royal North Shore samples (N=1047), the original association with fat deposition was not found. The minor allele frequency (T of the T/C SNP) was 0.16 and the p-value was 0.183 (QTDT analysis).

[0256] Haplotype analysis was performed using a program known as QPDT (Martin et al., *Amer. J. Human Genetics,* 67: 146-54 (2000)), which utilizes the EM algorithm (Dempster et al., *J. Royal Statistical Soc., B*39: 1-38 (1977)). The program was utilized to assign haplotypes based on likelihood of maximization. Table 7 shows possible haplotypes for four SNPs in the P2X4 gene and estimated frequencies for each.

TABLE 7

	Nucleotide Position in SEQ ID NO: 1				
Allele	11030	15847	21708	22713	Frequency
H1	T	T	A	T	0.54949
H2	A	C	A	T	0.21584

TABLE 7-continued

	Nucleotide Position in SEQ ID NO: 1				
Allele	11030	15847	21708	22713	Frequency
H3	T	C	A	T	0.12152
H4	A	C	G	C	0.10964

[0257] Haplotype versus single position association analysis for the P2X4 gene suggested that the H1 haplotype and H3 haplotype were most significantly associated with leanness.

Example 3

Association of an Insertion in P2X4 with Fat Deposition

[0258] The genetic analysis described in Example 2 identified three SNPs associated with fat deposition. Further studies were conducted to identify new SNPs or other polymorphic variations associated with fat deposition in the first six exons of P2X4. Samples from 30 individuals of the study group were utilized in this study where 10 individuals were homozygous for the H1 haplotype in Table 7, 10 individuals were homozygous for the H2 haplotype, and 10 individuals were homozygous for the H4 haplotype. The haplotypes were chosen for several reasons. First, choosing individuals from the same haplotype would increase the probability of identifying a SNP in linkage disequilibrium with other SNPs in the haplotype associated with fat deposition. Second, individuals homozygous for the H1 haplotype were characterized as having significantly decreased central fat deposits compared to individuals homozygous for the H2 and H4 haplotypes. Therefore, the three haplotypes represent two distinct groups of individuals, and therefore were used to determine differences in fat deposition for newly discovered SNPs. Exons 1-6 in the P2X4 nucleic acid were sequenced in the following manner.

[0259] A 50 μl PCR reaction was prepared to amplify a region to be sequenced. Typical reagent concentrations were 1× Qiagen PCR buffer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 200 nM forward primer, 200 nM reverse primer, 1 Unit HotStarTaq polymerase (Qiagen), and 25 ng genomic DNA. Forward and reverse primers used for the analysis are shown in Table 8. Primers were utilized in sequencing and PCR reactions, except pairs marked with an asterisk (*), which were utilized only in sequencing reactions, and pairs marked with a double asterisk (**), which were utilized only in PCR reactions.

TABLE 8

	Forward Primer			Reverse Primer		
Region in P2X4	Location in SEQ ID NO:1	Sequence	SEQ ID NO:	Location in SEQ ID NO:1	Sequence	SEQ ID NO:
Exon 1	3361—3344	GAGACGAGGTCTCGC TATGT		2655—2674	CCCAGGCTGTGG TTTTAA	
Exon 1*	2990—3008	GGGACCCAGACCGAC TAG		2984—3001	CAGTCCCCTAGT CGGTCTG	

TABLE 8-continued

	Forward Primer			Reverse Primer		
Region in P2X4	Location in SEQ ID NO:1	Sequence	SEQ ID NO:	Location in SEQ ID NO:1	Sequence	SEQ ID NO:
Exon 2	10314— 10333	TCCCCTAGAAGGTAC GTAGC		9855— 9874	CTGCCTGGATAC ACTCTGTT	
Exons 3&4	15107— 15125	GCGGTCAGTGTTTGA GTTG		14675— 14693	TGTGAAGGAATT GGGAAGA	
Exon 5	16074— 16092	TGTCTTCCTCCGATTC TAACT		15636— 15656	GTGGTGGGTGCC TATAATC	
Exon 6	22042— 22059	Not Used		Not Used	GCCTGAGTGCTG GTCTTC	
Exon 6	21705— 21722	AACATGGTTATGTGA GTATTTG		21336— 21357	CATGTCCTGGAA ACTGTG	
Exon 7*	None designed	CCTCAAGTCGTGCAT TTATG		21623— 21642	None designed	
Promoter Sequence	1275— 1295	CCAAAGTGCTGGGAT TAAA		464— 482	GTCTCTAACCGC TCTCATCTG	
Promoter Sequence	1994— 2014	CCAGGAAATGAGCTA AAAATG		1210— 1230	CCACTGAGAAA ATGCTATCAC	
Promoter Sequence	2772— 2790	CACATGCGGAAAATT GTATTC		1918— 1938	CGAAAGGCTCCT TCAAAAG	
Promoter* Sequence	918— 936	GAGGATTTAGCAATG GAGTTT		857— 877	TATCTGCCCAAA TGGAGTT	
Promoter* Sequence	1642— 1659	GCCTGTAATCCCAAC ACTT		1603—	CTCCTGACCTCC CATGAT	
Promoter* Sequence	2352— 2370	TACCCCTCACTTGAA CTCTG		2322— 2341	GCGAAGAGAGA GTGACCTC	
Exon 1**	2900— 2917	CGAGTTCCATGCCCC CTAGAG		2820— 2840	CCCGGAAGGAG GCAGCAC	
Exon 12	27042— 27061	CCTGGGACCAACTTG AGA		26283— 26300	GAGCCGGAATA GTGGATTTA	
Exons 9 to 11	26003— 26022	TGTGCCAGCTCCACT CTAAC		25200— 25219	GGCTCCCACACC CAAGAGAT	
Exons 9 to 11*	25662— 25680	TCATCCCCACTATGA TCAACAT		25499— 25520	GGCCAAGGGAA AGAAGGTA	
Exon 12*	26744— 26763	CCAGGAGGTCAGCAG TCTGT		26606— 26625	CCTGGAGAGAG CTGGAGACA	
Exon 1**	Same as 2820—2840, but with 5' SP6 sequence	CAGTAATACGACTCA CTATAGGGAGAAGGC TCGAGTTCCATGCCC CCTAGAG		Same as 2820—2840, but with 5' T7 sequence	CGATTTAGGAGA CACTATAGAAGA GCCCGGAAGGA GGCAGCAC	

[0260] 5 μl of each PCR product was analyzed on an agarose gel to confirm the size and purity of the amplified product. PCR products were purified using a Mo Bio Ultra Clean PCR kit. 45 μl of PCR product was mixed with 225 μl of Spin Bind buffer. The mixture was loaded onto a filter column and centrifuged for 30 seconds at 13,000 rpm. The solution was decanted. 300 μl of Spin Clean buffer was added to the filter column and centrifuged for 30 seconds at 13,000 rpm. The solution was decanted and the column was centrifuged again at 13,000 rpm for one minute. The filter column was placed in a new tube and 30 μl of water was added to the center of the filter column. The column was centrifuged for one minute at 13,000 rpm to elute DNA from column.

[0261] Sequencing reactions were performed with the following components: 2 μl of purified PCR product, 1 μl of sequencing primer at a 3.2 μM concentration, and 2 μl of

Applied Biosystems BigDye version 3.0 Ready Mix. The sequencing reaction was cycled with the following temperature program: (a) 96° C. for 15 seconds; (b) 55° C. for 15 seconds; (c) 60° C. for 4 minutes; (d) repeat steps (a) through (c) 24 times; and (d) 4° C. for completing the reaction.

[0262] Excess dye terminators were removed by precipitating the reactions. 25 μl of a sodium acetate-ethanol solution (1 μl of 3M sodium acetate, pH 4.6 per 25 μl of 100% ethanol) was added to each sequencing reaction. Reactions were centrifuged at 4,000 rpm for 30 minutes and the solution was decanted.

[0263] 100 μl of 70% ethanol was added to each reaction and centrifuged at 4,000 rpm for 10 minutes and the solution was decanted. 4 μl of loading buffer was added to each sample, and samples were vortexed. Loading buffer was a mixture of 1 part Blue Dextran/EDTA to 5 parts formamide. Samples were heated to 95° C. for 5 minutes to denature, then cooled to 4° C. 1.7 μl was loaded onto a 5% acrylamide, 6M urea gel and run for 7 hours on an Applied Biosystems Model 377 sequencer.

[0264] Prior to receiving nucleic acid samples for the 30 individuals, 17 random samples were chosen for sequencing of P2X4 exons 1 to 6. A seven-base pair insertion located in exon 1 within positions 2878 to 2891 of SEQ ID NO:1 was a newly identified polymorphism. In a majority of samples the sequence spanning positions 2878 to 2891 in SEQ ID NO:1 had two GGGCCCC repeats such that the sequence was 5'-GGGCCCCGGGCCCC-3' (SEQ ID NO:). A GGGCCCC insertion was identified in a minority of the samples, thereby yielding three GGGCCCC repeats in individuals having the insert instead of the two repeats shown in SEQ ID NO:1, which led to the sequence 5'-GGGCCCCGGGCCCCGGGCCCC-3' (SEQ ID NO:) instead of 5'-GGGCCCCGGGCCCC-3' (SEQ ID NO:). The frequency of the insertion was 2.9% in the 17 samples, which corresponds to one insertion out of 34 alleles. In the group of 30 samples, the insertion allele frequency was 32.8% (19 out of 58 alleles, sequences were not interpreted for one individual).

[0265] Within the 30 samples, eight individuals were homozygous for the insertion and three were heterozygous for the insertion. Central fat values for the homozygous insertion individual versus homozygous deletion individuals were compared to one another using the Student's T-test, and it was determined that individuals homozygous for the insertion had a significantly higher prevalence of central fat deposition (p=0.0064). Specifically, individuals homozygous for the insertion (N=9) possessed an average central fat value of 1993.42 grams (SEM=178.04), whereas individuals homozygous for the deletion (N=18) had an average central fat value of 892.51 grams (SEM=184.21). Thus, the percent increase in central fat deposition between the two groups was 53.66%, significantly showing the importance of the insert for diagnosing a predisposition to fat deposition.

[0266] This result, however, did not replicate when the insertion/deletion polymorphism was genotyped in a larger population. The insertion/deletion polymorphism located at position 2878 was originally genotyped over 30 dizygotic and monozygotic twins from the St. Thomas and Royal North Shore samples described in Example 1, and was found to be associated with fat deposition as described above. Upon genotype analysis over the remaining St. Thomas and Royal North Shore samples (N=1011), the original association with fat deposition was not found. The minor allele frequency (Insertion of the Insertion/Deletion polymorphism) was 0.09 and the p-value was 0.252 (QTDT analysis).

Example 4

NIDDM Sample Selection

[0267] Pooling Strategies

[0268] Samples were placed into one of four groups based on disease status. The four groups were female case samples, female control samples, male case samples, and male control samples. A select set of samples from each group were utilized to generate pools, and one pool was created for each group. Each individual sample in a pool was represented by an equal amount of genomic DNA. For example, where 25 ng of genomic DNA was utilized in each PCR reaction and there were 200 individuals in each pool, each individual would provide 125 μg of genomic DNA. Inclusion or exclusion of samples for a pool was based upon the following criteria and detailed in the tables below. Selection criteria for the study described herein included patient ethnicity and diagnosis with NIDDM. Other phenotypic data collected included GAD antibody concentration, HbAlc concentration, body mass (BMI), patient age, date of primary diagnosis, age of individual as of primary diagnosis (See Table 9A below). Cases with elevated GAD antibody titers and low age of diagnosis were excluded to increase the homogeneity of the diabetes sample in terms of underlying pathogenesis. Controls with elevated HbAlc were excluded to remove any undiagnosed diabetics. Control samples were derived from non-diabetic individuals with no family history of NIDDM. Secondary phenotypes were also measured in the diabetic cases, phenotypes such as HDL, LDL, triglycerides, insulin, C-peptide, nephropathy status, neuropathy status, to name a few, which will allow secondary analysis of the cases the be performed in order to elucidate the potential pathway of the disease gene.

TABLE 9A

Exclusion Criteria	No. of individuals fulfilling exclusion criteria	Actual no. of samples excluded after each stage	No. of samples remaining
ALL SAMPLES Lack of availability of sample	34	34	1591
ALL SAMPLES Non-German ethnicity	261	239	1352
CASES GAD Ab >0.9	102	84	1268
CONTROLS HbA1c >6 or BMI >40	21	20	1248
CASES age <90	17	6	1242
CASES Age of Diagnosis <35,	150	203	1039
CONTROLS Family History of Diabetes	170		
CONTROLS Age-matching to case pool	43	43	996

[0269] The selection process yielded the pools set forth in Table 9B, which were used in the studies described herein. It should be noted that the diabetic cases are borderline obese and the controls are slightly overweight on average, where the World Health Organization classifies a BMI less than 25 as lean, and a BMI greater than 30 as obese.

TABLE 9B

	Female case	Female control	Male cases	Male control
Pool size (Number)	244	244	254	254
Pool Criteria (ex: case/control)	case	control	case	control
Mean Age (ex: years)	52.49	49.02	49.78	50.57
BMI (kg/m^2)	31.61	25.23	30.02	25.93

Example 5

Association of Polymorphic Variations with NIDDM

[0270] Blood samples were taken from individuals in the study population described in Example 4. Genomic DNA was extracted from these blood samples using standard techniques (BACC2 DNA extraction kit (Nucleon Biosciences)) and subjected to analysis. Based upon the the coexistence of all of the following or differing combinations of central fat, hypertension, glucose intolerance, dyslipidemia (elevated triglycerides and low HDL cholesterol), and impaired insulin stimulated glucose uptake ("insulin resistance") in a common disorder referred to as syndrome X, it was postulated that polymorphic variants associated with the development of central obesity would also be associated with NIDDM.

[0271] The SNP at position 15847 of SEQ ID NO:1 was also allelotyped and genotyped in NIDDM and non-NIDDM patients from the pool described above (see Example 4), using the same assay described in Example 2, (i.e., same primers provided in Tables 4 and 5).

[0272] Genotype Analysis

[0273] Table 10A and 10B show the allelotyping and genotyping results, respectively, for males and females. Statistical significance of each association was determined by the computer program QTDT (Abecasis et al., *Amer. J. Human Genetics,* 66: 279-292 (2000)). Transmission/disequilibrium tests (TDT), population stratification tests, and total association tests were utilized in this computer program. Because population stratification could have lead to spurious results from total association testing, association testing was considered only when there was no evidence of population stratification.

TABLE 10A

Allelotyping Results

SNP Position 15847 in SEQ ID NO: 1	Allele Frequency Cases	Allele Frequency Controls	Statistical Significance
Females	C = 0.500 **T** = 0.500	C = 0.583 T = 0.417	p = 0.017
Males	C = 0.535 T = 0.465	C = 0.501 T = 0.499	p = 0.341

[0274]

TABLE 10B

Genotyping Results

SNP Position 15847 in SEQ ID NO: 1	Allele Frequency Cases	Allele Frequency Controls	Statistical Significance
Females	C = 0.438 **T** = 0.562	C = 0.504 T = 0.496	p = 0.038
Males	C = 0.472 **T** = 0.528	C = 0.440 T = 0.560	p = 0.311

[0275] Both allelotyping and genotyping analysis revealed that a thymine at position 15847 of SEQ ID NO: 1 is associated with NIDDM in females. Similarly, a thymine at the same position was also found to be associated with central obesity (see Example 2). Therefore, the data demonstrates this SNP serves as a marker for an increased risk of developing obesity or diabetes either separately or together as part of a greater metabolic syndrome.

Example 6

Sequencing of P2X4, Identifying Novel SNPs and Determining Linkage Disequilibrium (LD)

[0276] The P2X4 gene was sequenced using ABI's Dye-Terminator Cycle Sequencing. The sequencing of the amplified DNA was carried out on ABI 377 sequencers. The sequences of the amplification products were determined using automated dideoxy terminator sequencing reactions with a dye terminator cycle sequencing protocol. The sequence data were further evaluated to detect the presence of polymorphic variants within the amplified fragments. The polymorphism search was based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position. These polymorphisms are described in Table 11 below.

TABLE 11

P2X4. consensus. seq. location	Gene.location	Individual.ID	Relative.to. rs1044249	Position.in. SEQ ID NO: 1
PLA2	exon3	rs5637		
P2RX7	intron7	rs208307	127801008	
P2RX7	exon11	rs1718119	127812223	
P2RX7	exon13	rs1621388	127819683	
P2RX7	3' downstream	rs2686384	127822211	
P2X4	promoter	P2RX4-Promoter1075	127841528	1075
P2X4	promoter	P2RX4-Promoter2383	127842836	2383

TABLE 11-continued

P2X4. consensus. seq. location	Gene.location	Individual.ID	Relative.to. rs1044249	Position.in. SEQ ID NO: 1
P2X4	promoter	P2RX4-Promoter2494	127842947	2494
P2X4	exon1	InDel	127843331	2878
P2X4	exon1	rs1044249	127843495	3042
P2X4	intron1	P2RX4-3191	127843644	3191
P2X4	intron2	SQNM-rs1182945	127851483	11030
P2X4	intron2	GEMINI-rs1182945	127851483	11030
P2X4	intron2	P2RX4-14744	127855197	14744
P2X4	exon4	P2RX4-14977	127855430	14977
P2X4	intron4	P2RX4-15046	127855499	15046
P2X4	intron4	P2RX4-rs1629287	127855512	15059
P2X4	exon5	rs1044251	127856300	15847
P2X4	intron5	rs725135	127857791	17338
P2X4	intron6	P2RX4-21538	127861991	21538
P2X4	exon7	P2RX4-21630	127862083	21630
P2X4	exon7	rs25644	127862161	21708
P2X4	intron8	rs1151880	127863166	22713
P2X4	intron8	P2RX4-25264	127865717	25264
P2X4	exon9	P2RX4-25346	127865799	25346
P2X4	intron10	P2RX4-25798	127866251	25798
P2X4	intron10	P2RX4-25864	127866317	25864
P2X4	intron11	P2RX4-26327	127866780	26327
P2X4	intron11	P2RX4-26334	127866787	26334
P2X4	intron11	P2RX4-26343	127866796	26343
P2X4	intron11	P2RX4-26360	127866813	26360
P2X4	Exon12	P2RX4-26739	127867192	26739
P2X4	Intron12	P2RX4-27043	127867496	27043
P2X4/ CAMKK2	3' downstream/ 3'-UTR	rs1653586	127871244	
P2X4/ CAMKK2	3' downstream/ 3'-UTR	rs1653589	127873206	
P2X4/ CAMKK2	3' downstream/ 3'-UTR	rs1063843	127877354	
CAMKK2	intron	rs2567983	127887358	
CAMKK2	intron	rs1718120	127902127	
CAMKK2	exon1	rs891780	127907828	

[0277] Some of the SNPs provided in Table were genotyped in approximately 1100 individuals from the population described in Example 1. The genotype frequencies are provided in Table 12 below.

TABLE 12

Individual.ID	Genotype Frequencies
rs5637	AA (28, 3%), AG (272, 26%), GG (760, 72%)
rs208307	CC (105, 10%), CG (423, 42%), GG (475, 47%)

TABLE 12-continued

Individual.ID	Genotype Frequencies
rs1718119	C (355, 35%), CT (514, 50%), T (157, 15%)
rs1621388	C (338, 35%), T (146, 15%), TC (471, 49%)
rs2686384	A (159, 15%), AG (518, 50%), G (354, 34%)
P2RX4-Promoter1075	A (27, 3%), G (788, 74%), G.A (252, 24%)
P2RX4-Promoter2383	A (21, 2%), A.C (233, 22%), C (797, 76%)
P2RX4-Promoter2494	A (415, 40%), G (136, 13%), G.A (494, 47%)
InDel	DD (833, 83%), ID (157, 16%), II (11, 1%)
rs1044249	
P2RX4-3191	G (736, 80%), G.T (137, 15%), T (49, 5%)
SQNM-rs1182945	A (461, 45%), AT (441, 43%), T (112, 11%)
GEMINI-rs1182945	AA (115, 11%), AT (455, 43%), TT (479, 46%)
P2RX4-14744	C (742, 71%), C.T (284, 27%), T (21, 2%)
P2RX4-14977	A (3, 0%), A.G (106, 10%), G (943, 90%)
P2RX4-15046	C (2, 0%), C.T (102, 10%), T (936, 90%)
P2RX4-rs1629287	A (117, 11%), G (466, 45%), G.A (453, 44%)
rs1044251	C (200, 20%), T (317, 32%), T/C (472, 48%)
rs725135	A (657, 66%), A/G (215, 22%), G (117, 12%)
P2RX4-21538	C (826, 81%), T (13, 1%), T.C (184, 18%)
P2RX4-21630	A (790, 80%), A/G (187, 19%), G (12, 1%)
rs25644	
rs1151880	C (12, 1%), C/T (185, 19%), T (792, 80%)
P2RX4-25264	–(16, 2%), T (839, 80%), T. –(189, 18%)
P2RX4-25346	A (1010, 96%), GA (37, 4%)
P2RX4-25798	C (17, 2%), G (784, 75%), GC (245, 23%)
P2RX4-25864	A (937, 90%), A.G (100, 10%), G (2, 0%)
P2RX4-26327	
P2RX4-26334	A (11, 1%), C (818, 78%), C.A (215, 21%)
P2RX4-26343	C (929, 89%), C.T (114, 11%), T (1, 0%)
P2RX4-26360	
P2RX4-26739	
P2RX4-27043	C (464, 44%), C.G (454, 44%), G (125, 12%)
rs1653586	G (834, 91%), T (2, 0%), TG (76, 8%)
rs1653589	C (903, 90%), T (3, 0%), TC (96, 10%)
rs1063843	C (623, 65%), T (35, 4%), TC (307, 32%)
rs2567983	
rs1718120	G (271, 26%), GT (504, 48%), T (270, 26%)
rs891780	

[0278] The extent of linkage disequilibrium (LD) between each pair of SNPs provided above was estimated as the difference between the observed two locus haplotype frequency using the major alleles at each SNP and the product of the observed major allele frequencies. The disequilibrium between SNPs was expressed by a common standardized metric, D' ($D/\min(p_1q_2, p2q_1)$), where p, and q, were the minor allele frequencies at two SNPs, and p_2 and q_2 were the corresponding major allele frequencies. Significant deviation of this disequilibrium from zero was tested by the use of a chi-square goodness-of-fit test. The results of the LD analysis are provided in Table 13 and in **FIG. 6**.

TABLE 13

	Haplotype	Freq	Rel.Freq	Cum.Rel.Freq
1	GCUAGCADGATCGTGCACATTACAACCGCCG	128	0:058	0:058
2	GGCCGGCADGATCGTGTACATTAGACCGGCCT	119	0:054	0:112
3	GGCCGGCADGATCGTGTACATTAGACCCGCTT	115	0:053	0:165
4	GGCCGGCADTATCGTGTACATTAGACCGGCCG	94	0:043	0:208
5	GGCCGGCADGATCGTGTACATTAGACCCGCTG	82	0:037	0:245

TABLE 13-continued

	Haplotype	Freq	Rel.Freq	Cum.Rel.Freq
6	GGCCGGAGIGTATGTACGTGC-AGACCGGCCT	72	0:033	0:278
7	GGCCGGCADGATCGTGTACATTAGACCGGCCG	50	0:023	0:301
8	GGCCGGCADGATCGTGTACATTAGACTGGCCG	45	0:021	0:322
9	GCTTAGCADGATCGTGTACATTAGACCGGCCT	44	0:020	0:342
10	GGGCGGCAIGTATGTGCGTGC-AGACCCGCCT	43	0:020	0:362
11	GCTTAGCGDGTATACACGCAUAGGCCCTTCG	38	0:017	0:379
12	GCTTAGCADGATCGTGTACATTAGACCCGCTT	37	0:017	0:396
13	GGTTAACGDGTACGTACACATTAGACCCGCCT	34	0:016	0:412
14	GCCCGGCADGATCGTGTACATTAGACCGGCCT	29	0:013	0:425
15	AGCCGGCADGATCGTGTACATTAGACCGGCCT	27	0:012	0:437
16	ACCCGGCADGATCGTGTACATTAGACCCGCTT	27	0:012	0:449
17	GGCCGGCADGATCGTGTACATTAGACTGGCCT	26	0:012	0:461
18	AGCCGGCADTATCGTGTACATTAGACCGGCCG	24	0:011	0:472
19	GCTTAACGDGTACGTACACATTAGACCCGCCT	23	0:011	0:483
20	GGTTAGCGDGTATACACGCATTAGGCCCTTCG	21	0:010	0:493
21	GGCCGGCADGATCGTGTACATTAGACCCGCCT	18	0:008	0:501
22	GCTTAGCGDGATCGTGCACATTACAACCGCCG	18	0:008	0:509
23	GGCCGGCADTATCGTGTACATTAGACCCGCTT	17	0:008	0:517
24	AGCCGGCADGATCGTGTACATTAGACCCGCTT	17	0:008	0:525
25	ACTTAGCADGATCGTGCACATTACAACCGCCG	17	0:008	0:533
26	GGTTAACGDGTACGTACGCATTAGACCCGCCG	16	0:007	0:540
27	GGTTAGCADGATCGTGTACATTAGACCGGCCT	15	0:007	0:547
28	GGTTAACGDGTACGTACACATTAGACCCGCCG	15	0:007	0:554
29	GGCCGGCADTATCGTGTACATTAGACCGGCCT	15	0:007	0:561
30	GGTTAACGDGTACGTACGCATTAGACCGGCCT	14	0:006	0:567
31	GGCCGGAGDTATCGTATACATTAGACGGGCCG	14	0:006	0:573
32	GGCCGACGDGTACGTACACATTAGACCCGCCG	14	0:006	0:579
33	ACTTAGCADGATCGTGTACATTAGACCGGCCT	14	0:006	0:S85
34	GCTTAGAGDGTAGGTACGCATTACACCCGGCG	13	0:006	0:591
35	GCTTAGAGDGTACGTACACATTACACCCGCCG	13	0:006	0:S97
36	AGTTAACGDGTACGTACGCATTAGACCCGCCT	13	0:006	0:603
37	GGCCGGAGIGTATGTACGTGG-AGACCCGCCG	12	0:005	0:608
38	GGCCGACGDGTACGTACACATTGGAGCCGCCG	12	0:005	0:613
39	GGCCGACGDGATCGTATACATTAGACCGGCCT	12	0:005	0:618
40	GGTTAGCADGATCGTGCACATTACAAGCGCCG	11	0:005	0:623
41	GGGCGACGDGATCGTATACATTAGACCCGCTT	11	0:005	0:628
42	GCTTAGCADGTATACGCGCATTAGGCCCTTCG	11	0:005	0:633

TABLE 13-continued

	Haplotype	Freq	Rel.Freq	Cum.Rel.Freq
43	GCTTAGCADGATCGTGCACATTACAACCGCCT	11	0:005	0:638
44	AGTTAGCADGATCGTGTACATTAGACCGGCCT	11	0:005	0:643
45	AGCCGGCADTATCGTGTACATTAGACCGGCCT	11	0:005	0:648
46	AGCCGGCADGATCGTGTACATTAGACCCGCTG	11	0:005	0:653
47	GGCCGGAGDGATCGTATACATTAGACCGGCGG	10	0:005	0:658
48	GGCCGACGDGATCGTATACATTAGACCGGCCG	10	0:005	0:663
49	GCTTAACGDGTACGTACGCATTAGACCCGCCT	10	0:005	0:668
50	GCCCGGCADGATCGTGTACATTAGACCCGCTT	10	0:005	0:673
51	GCTTAACGDGATCGTATACATTAGACCCGCTT	9	0:004	0:677
52	AGCCGGAGIGTATGTACGTGC-AGACCCGCCT	9	0:004	0:681
53	ACTTAGCADGATCGTGTACATTAGACCCGCTT	9	0:004	0:685
54	ACTTAGAGDGATCGTACACATTACAACCGCCG	9	0:004	0:689
55	GGCCGGCADGATCGTGTACATTAGACCCGCCG	8	0:004	0:693
56	GGCCGACGIGTATGTACGTGC-AGACCCGCCT	8	0:004	0:697
57	GGCCGACGDTATCGTATACATTAGACCGGCGG	8	0:004	0:701
58	AGTTAGCADGTACGTGCGCATTAGACCCGCCT	8	0:004	0:705
59	AGTTAACGDGTACGTACACATTAGACCCGCCG	8	0:004	0:709
60	ACTTAGAGDGTACGTACGGATTACACCCGCCG	8	0:004	0:713
61	GGTTAGCAIGTATGTGCGTGC-AGACCCGCCG	7	0:003	0:716
62	GGCCGGCGDGATCGTATACATTAGACCGGCCT	7	0:003	0:719
63	GGCCGGCADGTACGTGCACATTAGACCCGCTG	7	0:003	0:722
64	GGCCGGAGIGTACGTACGCGCTAGACCCGCCT	7	0:003	0:725
65	GGCCGGAGDGATCGTATACATTAGACCGGCCT	7	0:003	0:728
66	GGCCGACGDGTACGTACGCATTAGACCCGCCG	7	0:003	0:731
67	GCTTAGCADGATCGTGTACATTAGACGGGCCG	7	0:003	0:734
68	GCTTAACGDGATCGTACACATTACAACCGCCG	7	0:003	0:737
69	AGTTAGCADGTACGTGCACATTAGACCCGCCG	7	0:003	0:740
70	ACCCGGCADGATCGTGTACATTAGACCGGGCT	7	0:003	0:743
71	GGTTAGCADGTATACGCGCATTAGGCCCTTCG	6	0:003	0:746
72	GGTTAGCADGTACGTGCGCATTAGACCCGCCT	6	0:003	0:749
73	GGTTAGCADGTACGTGCACATTAGACCCGCCG	6	0:003	0:752
74	GGTTAGCADGATCGTGTACATTAGACCCGCTT	6	0:003	0:755
75	GGTTAGCADGATCGTGTACATTAGACCCGCCT	6	0:003	0:758
76	GGTTAGAGIGTATGTACGTGC-AGACCCGCCG	6	0:003	0:761
77	GGTTAGAGDGTACGTACACATTAGACCCGCCT	6	0:003	0:764
78	GGCCGGCADGTACGTGCACATTGGACCCGCCG	6	0:003	0:767
79	GGCCGGAGDGATCGTATACATTAGACCCGCTT	6	0:003	0:770

TABLE 13-continued

	Haplotype	Freq	Rel.Freq	Cum.Rel.Freq
80	GCTTAGCADGATCGTGTACATTAGAGTGGCCG	6	0:003	0:773
81	GCTTAGAGDGATCGTATACATTAGACCCGGTT	6	0:003	0:776
82	GCCCGGCADGATCGTGTACATTAGACTGGCCT	6	0:003	0:779
83	GCCCGGCADGATCGTGTACATTAGACCCGCTG	6	0:003	0:782
84	GGCCGGCADTATCGTGTACATTAGACCCGCTG	5	0:002	0:784
85	GGCCGACGDGTACGTACGCATTGGACCCGCCG	5	0:002	0:786
86	GCTTAGCGDGTACGTACGCATTAGACCCTTCG	5	0:002	0:788
87	GCTTAGCADGTACGTGCGCATTAGACCCGCCG	5	0:002	0:790
88	GCTTAAAGDGTACGTACGCATTAGACCCGCCG	5	0:002	0:792
89	GCTTAAAGDGTACGTACAGATTAGACCCGCGG	5	0:002	0:794
90	GGCCGGCGDGATCGTGTACATTAGACCGGCCT	5	0:002	0:796
91	GCCCGGCADGATCGTGTACATTAGAGCGGCCG	5	0:002	0:798
92	ACCCGGCADGATCGTGTACATTAGACCGGCCG	5	0:002	0:800
93	GGTTAGCGDGTACGTACGCATTAGACCGGCCT	4	0:002	0:802
94	GGTTAGCADGTACGTGTACATTAGACCCGCGT	4	0:002	0:804
95	GGTTAGCADGTACGTACGCATTAGACCCGCCT	4	0:002	0:806
96	GGTTAGAGDGTACGTGCACATTAGACCCGCCT	4	0:002	0:808
97	GGTTAACGDTTACGTACACATTAGACCCGCCG	4	0:002	0:810
98	GGCCGGCGDGTATACACGCATTAGGCCCTTCG	4	0:002	0:812
99	GGCCGGCGDGATCGTGTACATTAGACCCGCTT	4	0:002	0:814
100	GGCCGGCGDGATCGTGTACATTAGACCCGCCG	4	0:002	0:816
101	GGCCGGCAIGTATGTGGGTGC-AGACCCGCCG	4	0:002	0:818
102	GGCCGGCAIGATCGTGTACATTAGACCGGCCG	4	0:002	0:820
103	GGCCGGCADGTACGTACACATTAGACCCGCCG	4	0:002	0:822
104	GGCCGGCADGATTGTGTATAT-AGACCCGCGG	4	0:002	0:824
105	GGCCGGCADGATCGTGCACATTACAACCGCCG	4	0:002	0:826
106	GGCCGGAGITTACGTACGCGCTAGACCCGCCG	4	0:002	0:828
107	GGCCGGAGDGTATGTACGTGC-AGACCCGCCT	4	0:002	0:830
108	GCTTAGCADGTACGTGCGCATTAGACCCGCCT	4	0:002	0:832
109	GCTTAGCADGATCGTGTACATTAGACCCGCCT	4	0:002	0:834
110	GCTTAGAGDGTACGTACACATTAGACCCGCCT	4	0:002	0:836
111	GCTTAACGDGATCGTATACATTAGACCGGCCT	4	0:002	0:838
112	GCCCGGAGDGATCGTATACATTAGACCGGCCG	4	0:002	0:840
113	AGTTAGCADGTACGTGCACATTAGACCCGCTT	4	0:002	0:842
114	AGTTAGCADGATCGTGTACATTAGACCCGCCT	4	0:002	0:844
115	AGTTAACGDGTACGTACGCATTAGACCCGCCG	4	0:002	0:846
116	AGCCGGCADGATCGTGTACATTAGACCCGCCG	4	0:002	0:848

TABLE 13-continued

	Haplotype	Freq	Rel.Freq	Cum.Rel.Freq
117	AGCCGGAGDGATCGTATACATTAGACTGGCCG	4	0:002	0:850
118	GGTTAGCGDGTACGTACACATTAGACCCGCCT	3	0:001	0:851
119	GGTTAGCADGTACGTGCACATTAGACTCGCCG	3	0:001	0:852
120	GGTTAGCADGATCGTGCACATTAGACCCGCCG	3	0:001	0:853
121	GGTTAGAGIGTATGTACGTGC-AGACCCGCCT	3	0:001	0:854
122	GGTTAGAGDGTACGTACGCA1TAGACCCGCCT	3	0:001	0:855
123	GGTTAACGDGTATACACACATTAGGCCCGCCT	3	0:001	0:856
124	GGCCGGCGDGTACGTGCGCATTACACCCGCCG	3	0:001	0:857
125	GGCCGGCGDGATCGTGTACATTAGACCGGCCT	3	0:001	0:858
126	GGCCGGCGDGATCGTATACATTAGACCGGCCG	3	0:001	0:859
127	GGCCGGCAIGATTGTATATGC-AGACCCGCTG	3	0:001	0:860
128	GGGCGGCADTATCGTGTACATTAGACCGGCTG	3	0:001	0:861
129	GGCCGGCADGTATGTGCGTGC-AGACCCGCCT	3	0:001	0:862
130	GGCCGGCADGATCGTGTACATTGGACCCGCTG	3	0:001	0:863
131	GGCCGGCADGATCGTGTACATTACAACCGCCG	3	0:001	0:864
132	GGCCGGCADGATCGTGCACATTAGACCCGCTT	3	0:001	0:865
133	GGCCGGAGITTATGTACGTGC-AGACCCGCCT	3	0:001	0:866
134	GGCCGGAGIGTACGTACGCGGTACAACCGCCT	3	0:001	0:867
135	GGCCGGAGIGTACATATGCGCTAGACCGGCCT	3	0:001	0:868
136	GGCCGGAGDGATCGTATACATTAGACTGGCCG	3	0:001	0:869
137	GCTTAGCGDGTACGTACGCATTAGACCGTTCG	3	0:001	0:870
138	GCTTAGCGDGATCGTGTACATTAGACCGGCCG	3	0:001	0:871
139	GCTTAGCGDGATCGTACACATTACAACCGCCG	3	0:001	0:872
140	GCTTAGCADTTACGTGCACATTAGACCCGCCG	3	0:001	0:873
141	GCTTAGCADTATCGTGCACATTACAACCGCCG	3	0:001	0:874
142	GCTTAGCADGTACGTGCACATTACACCCGCCG	3	0:001	0:875
143	GCTTAGCADGATCGTGCACATTAGACTGGCCG	3	0:001	0:876
144	GCTTAAGGDTATCGTGTACATTAGACCGGCCT	3	0:001	0:877
145	GGTTAACGDGTATACACGCATTAGGCCCTTCG	3	0:001	0:878
146	GCCCGGCADTATCGTATACATTAGACCGGCCT	3	0:001	0:879
147	AGTTAGCGDTATCGTATACATTAGACCCGCCT	3	0:001	0:880
148	AGTTAGCAIGTATGTGCGTGC-AGACCCGCCT	3	0:001	0:881
149	AGTTAGCADGATCGTGGAGATTACAACGGCCG	3	0:001	0:882
150	AGTCAGCGDGATCGTATACATTAGACCCGCCT	3	0:001	0:883
151	AGTCAACGDGTATACACGCATTAGGCCCTTCG	3	0:001	0:884
152	AGCCGGCGDGTACGTGCGCATTAGAACCGCCG	3	0:001	0:885
153	AGCCGGCGDGATCGTATACATTAGACTGGCCG	3	0:001	0:886

TABLE 13-continued

	Haplotype	Freq	Rel.Freq	Cum.Rel.Freq
154	AGCCGACGDGTACGTACACATTGGACCCGCCG	3	0:001	0:887
155	ACTTAACGDGTACGTACACATTAGACCCGCCG	3	0:001	0:888
156	GGTTAGCGDGATCGTATACATTAGACTGGCCT	2	0:001	0:889
157	GGTTAGCADGATCGTGCGCATTACAACCGCCG	2	0:001	0:890
158	GGTTAGAGDGTACGTACACATTAGACCCGCCG	2	0:001	0:891
159	GGTTAACGDGTATGTACATAT-ACAACCGCCT	2	0:001	0:892
160	GGTTAACGDGTACGTACGCATTAGACCCGCTT	2	0:001	0:893
161	GGTTAACGDGAACGTACACATTACAACCGCCT	2	0:001	0:894
162	GGCTGGCADGATCGTGTACATTAGACCGGCCG	2	0:001	0:895
163	GGCCGGCGIGTATGTACGTGC-AGACTGGCCG	2	0:001	0:896
164	GGCCGGCGDTATCGTGTACATTAGACCGGCCG	2	0:001	0:897
165	GGCCGGCGDGTACGTACGCATTAGACCCGCTG	2	0:001	0:898
166	GGCCGGCGDGATCGTGTACATTAGACTGGCCG	2	0:001	0:899
167	GGCCGGCGDGATCGTGTACATTAGACCGGCCG	2	0:001	0:900
168	GGCCGGCADTTACGTGCACATTGGACCCGCCT	2	0:001	0:901
169	GGCCGGCADTATCGTGTGCATTAGACTGGCCG	2	0:001	0:902
170	GGCCGGCADTATCGTGTGCATTAGACCGGCCG	2	0:001	0:903
171	GGCCGGCADTATCGTGCACATTAGACCGGCCG	2	0:001	0:904
172	GGCCGGCADTATCGTGCACATTACAACCGCCT	2	0:001	0:905
173	GGCCGGCADGTACGTGCGCATTGGACCCGCCG	2	0:001	0:906
174	GGCCGGCADGTACGTGCGCATTAGACCCGCCG	2	0:001	0:907
175	GGCCGGCADGTACGTACGCATTAGACCCGCCG	2	0:001	0:908
176	GGCCGGCADGATTACGTACATTAGGCCCGCCT	2	0:001	0:909
177	GGCCGGCADGATTACGCACGTTAGGCCGTTCG	2	0:001	0:910
178	GGCCGGCADGATCGTGTACATTAGACCGGCTT	2	0:001	0:911
179	GGCCGGAGIGTTTGTACGTGC-AGACCCGCCT	2	0:001	0:912
180	GGCCGGAGIGTACGTACGCGCTAGACCGGCCT	2	0:001	0:913
181	GGCCGGAGDGTATGTACGTGC-AGACCCGCCG	2	0:001	0:914
182	GGCCGGAGDGTATGTACATGC-AGACCCGCCT	2	0:001	0:915
183	GGCCGGAGDGATCGTATACATTAGACCCGCTG	2	0:001	0:916
184	GGCCGGAADGAACGTGTACATTAGACTGGCCT	2	0:001	0:917
185	GGCCGACGITTATGTATGTGT-AGACCCGCCT	2	0:001	0:918
186	GGCCGACGDGTACGTACACATTAGACTGGCCG	2	0:001	0:919
187	GGCCGACADTATCGTACGCATTAGACCCGCCG	2	0:001	0:920
188	GCTTGGCAIGATCGTGTACATTAGACCCGCCG	2	0:001	0:921
189	GCTTAGCGDGTACGTACAGATTACACCCGCCG	2	0:001	0:922
190	GCTTAGCGDGATCGTGCACATTACACCCGCCG	2	0:001	0:923

TABLE 13-continued

	Haplotype	Freq	Rel.Freq	Cum.Rel.Freq
191	GCTTAGCAIGATTGTGCGTGC-ACAACCGCTT	2	0:001	0:924
192	GCTTAGCADGTACGTGCACATTAGACCCGCCT	2	0:001	0:925
193	GCTTAGCADGATTGTGCATAT-AGACCCGCCG	2	0:001	0:926
194	GCTTAGCADGATCGTGTACATTAGACCGGCTT	2	0:001	0:927
195	GCTTAGAGDGTACGTACGCATTAGACCCGCCG	2	0:001	0:928
196	GCTTAGAGDGTACGTACACATTAGACCCGCCG	2	0:001	0:929
197	GCTTAGAGDGATCGTATACATTAGACCGGCCT	2	0:001	0:930
198	GCWVAGAADTATCGTACAGATTACAACCGCCT	2	0:001	0:931
199	GCCCGGCGDGTTTACACGCATTAGGCCCTTCT	2	0:001	0:932
200	GCCCGGGGDGATCGTATACATTAGACCCGCTG	2	0:001	0:933
201	GGCCGGCADTATVGTGTATAT-AGACCCGCCT	2	0:001	0:934
202	GGCCGGAGIGTATGTACGTGC-AGACCCGCCG	2	0:001	0:935
203	GCCCGGAGIGTACGTACGCGCTAGACCGGCCT	2	0:001	0:936
204	GCCCGGAGDGATTGTGCGTGC-ACAACCGCCT	2	0:001	0:937
205	GCGCAGCADGATCGTGTACATTAGACCGGCCG	2	0:001	0:938
206	AGTTAGCADTATCGTGTAGATTAGACCGGCCG	2	0:001	0:939
207	AGTTAGAGDGTACGTACGCATTAGACCCGCCT	2	0:001	0:940
208	AGTTAACGIGTACGTACACATTAGACCCGCCG	2	0:001	0:941
209	AGCCGGCGDTTATGTATATAT-AGACCCGCCG	2	0:001	0:942
210	AGCCGGCADTATCGTGTGCATTAGACCGGCCT	2	0:001	0:943
211	AGCCGGCADGATTGTGCATAT-AGACCCGCCG	2	0:001	0:944
212	AGCCGGCADGATCGTGTACATTAGACTGGCCG	2	0:001	0:945
213	AGCCGGCADGATCGTATACATTAGACTGGCCG	2	0:001	0:946
214	AGCCGACGDGTACGTAGGCATTAGACCCGCTG	2	0:001	0:947
215	AGCCGACGDGTACGTACACATTAGACCCGCCG	2	0:001	0:948
216	AGTTAGCADTATCGTGTACATTAGACCCGCTT	2	0:001	0:949
217	ACTTAAGGDGAATACATGCATTAGGCGCTTCT	2	0:001	0:950
218	GGTTAGCGDGATCGTGCACATTACAACCGCCG	1	0:000	0:950
219	GGTTAGGGDGATCGTATACATTAGACCGGCCT	1	0:000	0:950
220	GGTTAGCGDGATCGTATACATTAGACCCGCCT	1	0:000	0:950
221	GGTTAGCAITTATGTGCGTGC-AGACCCGCCG	1	0:000	0:950
222	GGTTAGCADTATCGTGTACATTAGACCGGCCT	1	0:000	0:950
223	GGTTAGCADGATTGTGCACAT-AGAACCGCCG	1	0:000	0:950
224	GGTTAGCADGATCGTGTACATTAGACCCGCCG	1	0:000	0:950
225	GGTTAGCADGATCGTGCACATTACAACCGCCT	1	0:000	0:950
226	GGTTAGAGIGAATGTACGTGC-AGACCCGCCT	1	0:000	0:950
227	GGTTAGAGDGTATACACGCATTAGGCCCTTCG	1	0:000	0:950

TABLE 13-continued

	Haplotype	Freq	Rel.Freq	Cum.Rel.Freq
228	GGTTAGAGDGTACATACGCATTAGACCCGCCT	1	0:000	0:950
229	GGTTAACGDTTACGTACACATTAGACCGGCCG	1	0:000	0:950
230	GGTTAACGDGTATACACGCATTAGGCCCGGCT	1	0:000	0:950
231	GGTCGGCGDGATCGTGTACATTAGACCGGCCT	1	0:000	0:950
232	GGCTGGCAIGTATGTGCGTGC-AGACCCGCGT	1	0:000	0:950
233	GGCTGACGDGTTTGTATGCATTAGACCCGCTT	1	0:000	0:950
234	GGCTGACGDGTTTGTATACATTAGACCCGCTT	1	0:000	0:950
235	GGGCGGCGDTATCGTGCACATTAGACCGGCCG	1	0:000	0:950
236	GGCCGGCGDGTTGGTATACATTAGACCGGCCT	1	0:000	0:950
237	GGCGGGCGDGATCGTGTACATTAGACCCGCTG	1	0:000	0:950
238	GGCCGGCGDGATCGTGTACATTAGACCCGCCT	1	0:000	0:950
239	GGCCGGCGDGATCGTATACATTAGACCCGCCG	1	0:000	0:950
240	GGCGGGCGDGATCGTACACATTAGACCGGCCT	1	0:000	0:950
241	GGCCGGCAIGATTGTGTATGC-AGACCCGCTG	1	0:000	0:950
242	GGCCGGCADTATCGTGTACATTAGACTGGCCG	1	0:000	0:950
243	GGCCGGCADGTACGTGCACATTGGACCCGCTG	1	0:000	0:950
244	GGCCGGCADGTACGTGCACATTAGACCCGCCG	1	0:000	0:950
245	GGCGGGCADGATTGCGTATGC-AGACCCGTTG	1	0:000	0:950
246	GGCCGGCADGATCGTGTACATTACACCGGCCT	1	0:000	0:950
247	GGCCGGCADGATCGTGTACATTAGACCCGCTG	1	0:000	0:950
248	GGCCGGCADGATCGTGTACAT-AGGCCGGCCT	1	0:000	0:950
249	GGCCGGCADGATCGTATACAT-AGACGGGCTT	1	0:000	0:950
250	GGCCGGCADGAACGTGTACATTAGACCCGCTG	1	0:000	0:950
251	GGCCGGAGDGTTCATATAGATTAGACCCTCGT	1	0:000	0:950
252	GGCCGGAGDGTACATAGACATTAGAGCCGCCT	1	0:000	0:950
253	GGCCGGAGDGATCGTGTACATTAGAGCCGCTT	1	0:000	0:950
254	GGCCGACGDTTACGTACACATTGGAGCCGCCG	1	0:000	0:950
255	GGCCGACGDGTACGTACGGATTAGACCGGCGG	1	0:000	0:950
256	GGCCGACGDGATCGTATACATTAGACCCGTGT	1	0:000	0:950
257	GGCCGACGDGATCGTACACATTACAACCGCCG	1	0:000	0:950
258	GCTTGGCADGATCGTGTACATTAGACCCGCCG	1	0:000	0:950
259	GCTTGGAGDGATCGTATACATTAGACCCGCTT	1	0:000	0:950
260	GCTTAGCGDTTATAGACGCATTAGGCCCTTCG	1	0:000	0:950
261	GCTTAGCGDGTATATACGCATTAGGCCCTTCG	1	0:000	0:950
262	GCTTAGCGDGTATACGCGCATTAGGCCCTTCG	1	0:000	0:950
263	GCTTAGCGDGTACGTACGCATTAGACCCGCCG	1	0:000	0:950
264	GCTTAGCGDGATCGTGCACATTAGACGCGCCG	1	0:000	0:950

TABLE 13-continued

	Haplotype	Freq	Rel.Freq	Cum.Rel.Freq
265	GCTTAGCGDGATCGTGCACATTACAACCGTCG	1	0:000	0:950
266	GCTTAGCGDGATCGTATACATTACAACCGCCG	1	0:000	0:950
267	GCTTAGCAIGATTGTACGTGC-ACAACCGCCT	1	0:000	0:950
268	GCTTAGCADTATCGTGTACATTAGACCGGCCG	1	0:000	0:950
269	GCTTAGCADGTATACGTGCATTAGGCCCTTCG	1	0:000	0:950
270	GCTTAGCADGTACGTGCACATTAGACCCGCCG	1	0:000	0:950
271	GCTTAGCADGATCGTGTACATTAGACCCGCTG	1	0:000	0:950
272	GCTTAGCADGATCGTGCGCATTACAACCGCCG	1	0:000	0:950
273	GCTTAGCADGATCGTACACATTACAACCGCCG	1	0:000	0:950
274	GCTTAGCADGAACGTGCACATTACAACCGCCG	1	0:000	0:950
275	GCTTAGAGDGTACGTATACATTAGACCCGCCT	1	0:000	0:950
276	GCTTAGAGDGTACGTACACGTTACACCCGCCG	1	0:000	0:950
277	GCTTAGAGDGTACGTACACATTAGACCGGCCT	1	0:000	0:950
278	GCTTAGAGDGTACATACGCATTAGACCCGCCT	1	0:000	0:950
279	GCTTAGAGDGATTACACACATTAGGCCGGCCG	1	0:000	0:950
280	GCTTAGAGDGATTACACACATTAGGCCCGCCG	1	0:000	0:950
281	GCTTAGAAIGATTGTACGTGC-ACACCCGCCT	1	0:000	0:950
282	GCTTAGAADGATCGTGTACATTAGACCGGCCT	1	0:000	0:950
283	GCTTAACGDTTACGTACACATTAGACCCGCCT	1	0:000	0:950
284	GCTTAACGDGATCGTATACATTAGACCCGCTG	1	0:000	0:950
285	GCTTAACGDGATCGTATACATTAGACCCGCCG	1	0:000	0:950
286	GC1TAACGDGATGGTAGACATTAGACCCGCCG	1	0:000	0:950
287	GCTCAGCGDGATCGTGTACATTAGACCGGCCG	1	0:000	0:950
288	GCCTAGCGDGTATACGCGCATTAGGCCCTTTG	1	0:000	0:950
289	GCCTAGCADGATCGTGCACATTAGAACCGCCG	1	0:000	0:950
290	GCCCGGCGDTTATACACGCATTAGGCCCTTCG	1	0:000	0:950
291	GCCCGGCGDGTATACACGCATTAGGCCCTTCG	1	0:000	0:950
292	GCCCGGCADGTACGTACACATTAGACCCGCCG	1	0:000	0:950
293	GGCCGGCADGATTACGTACATTAGGCCCGCCT	1	0:000	0:950
294	GCCCGGAGIGATTGTACGTGC-ACACCCGCCT	1	0:000	0:950
295	GCCCGGAGDGATTGTGCATGC-ACAACCGCCT	1	0:000	0:950
296	GCCCGACGDGTAGGTACACATTGGACCCGCCG	1	0:000	0:950
297	GCCCGACGDGTACGTACACATTAGACCCGCCG	1	0:000	0:950
298	GCCCAGCADGTTCGTATATATTAGACCCGTTT	1	0:000	0:950
299	AGTTAGCADGTATGTGCGTGC-AGACGCGCCT	1	0:000	0:950
300	AGTTAACGIGTACGTATACATTAGACCCGCCG	1	0:000	0:950
301	AGTTAACGDTATCGTATACATTAGACCCGCCT	1	0:000	0:950

TABLE 13-continued

	Haplotype	Freq	Rel.Freq	Cum.Rel.Freq
302	AGCCGGCGIGTATGTACGTGC-AGACCCGGCG	1	0:000	0:950
303	AGCGGGCGDGATCGTGTACATTAGACCCGCTT	1	0:000	0:950
304	AGCCGGCADTATCGTGTACATTAGACCCGCTT	1	0:000	0:950
305	AGCCGGCADTATCGTGTACATTAGACCCGCCG	1	0:000	0:950
306	AGCGGGCADTATCGTGTAGATTACACCGGCCT	1	0:000	0:950
307	AGCCGGCADGATTGTGTATAT-AGACCCGCCG	1	0:000	0:950
308	AGCCGGCADGATGGTGCACATTAGACCCGCTT	1	0:000	0:950
309	AGCCGGAGIGTATGTACGTGC-AGACCGGCCG	1	0:000	0:950
310	AGCGGACGIGTATGTACGTGC-AGACCCGCCT	1	0:000	0:950
311	AGCCGACGDTATCGTATACATTAGACTGGCCG	1	0:000	0:950
312	AGCCGACGDGTACGTACACATTAGACCGGCCG	1	0:000	0:950
313	AGCCGACADGTACGTACGCATTAGACCCGCTG	1	0:000	0:950
314	ACTTAGCADGATCGTGTACATTAGACCGGCCG	1	0:000	0:950
315	ACTTAGAGDGTACGTACGCAT-ACACGCGCCG	1	0:000	0:950
316	ACTTAACADGATCGTGCACATTACAACCGCCT	1	0:000	0:950
317	ACTTAAAGDTATCGTACAGATTACAACGGCCG	1	0:000	0:950
318	ACCCGGCGDGTACGTACGCATTGGACCCGCCG	1	0:000	0:950

[0279] **FIG. 6** demonstrates that polymorphic variants in a region spanning from about position 11000 to about position 25300 in SEQ ID NO:1 are in linkage disequilibrium. Polymorphic variants occur at positions 11030, 14744, 14977, 15046, 15059, 15847, 17338, 21538, 21630, 21708, 22713, and 25264,

Example 7

P2X4 Tissue Expression Profiles

[0280] Expression levels of P2X4 were determined in tissues of Israeli sand rats (*P. obesus*) by detecting RNA transcribed from the P2X4 gene. P2X4 expression levels are depicted graphically in **FIG. 4** for lean sand rats. P2X4 was ubiquitously expressed among the tissues and expression levels were comparatively higher in small intestine, cerebellum, cortex, large intestine, subscapular fat, lung, and kidney. Expression was also detected in the hypothalmus and liver.

[0281] For the same tissues set forth in **FIG. 4**, P2X4 expression was also monitored among sand rats in three different groups that were exposed to different diets. Specifically, Group A rats are lean animals, Group B rats are hyperinsulinemic and obese, and Group C rats are hyperinsulinemic, hyperglycemic, and obese (diabetic). The mice are classified into Groups A, B, or C at 16 weeks of age on the basis of body weight, blood glucose, and blood plasma insulin measurements made at 8, 12, and 16 weeks of age. Rats of Groups A, B, and C were identified within the same litter. Sand rats used in the studies were subjected to various stresses to determine whether a particular stimulus caused and changed the expression level of P2X4. In particular, two studies were performed. In one study, expression levels of P2X4 were compared in rats that were fed versus rats that fasted 24 hours prior to tissue harvest. In a second study, P2X4 expression levels were compared for rats fed 70% of normal food intake for two weeks prior to tissue harvest versus control rats. In these studies, a significant difference in P2X4 expression levels in the pancreas was identified between Group C energy restricted rats and control rats with statistical certainty (p=0.004). Also, a significant difference in pancreatic expression of P2X4 was identified between Group C energy restricted rats and Group A control rats (p=0.019). Further, where rats in Groups A, B, and C were combined into energy restricted groups and control groups, a significant difference in pancreatic P2X4 expression was identified (p=0.037). In all three of these associations, increased pancreatic P2X4 expression was observed in energy restricted rats.

[0282] Gene expression was quantified using a TaqMan™ PCR system (ABI Prism™ 7700 Sequence Detection System, Perkin-Elmer Applied Biosystems, Norwalk, USA) and was determined relative to an endogenous control gene, cyclophilin. cDNA was synthesized by subjecting one microgram of total RNA to a reverse transcription reaction using SuperScript II RNase H-Reverse Transcriptase (Invitrogen) according to manufacturer's instructions (see http address www.invitrogen.com/Content/World/11904018.pdf). In this reverse transcriptase PCR (RT-PCR) procedure, the following contents were added to a nuclease-free microcentrifuge tube: 1 μl Oligo (dT)12-18 (500 μg/ml);

1 μg total RNA; 1 μl 10 mM dNTP mix (10 mM each dATP, dGTP, dCTP and dTTP at neutral pH); sterile, distilled water to 12 μl. The mixture was heated to 65° C. for 5 minutes and quickly chilled on ice for at least 2 minutes.

[0283] Contents of the tube were collected by brief centrifugation and the following were added to complete a 20-μl reaction volume: 2 μl 10× First-Strand Buffer; 4 μl 25 mM $MgCl_2$; 2 μl 0.1 M DTT; 1 μl RNaseOUT Recombinant Ribonuclease Inhibitor (40 units/μl); 1 μl (200 units) of and SUPERSCRIPT II. The mixture was incubated at 45° C. for 50 minutes and then the reaction was inactivated by heating at 70° C. for 15 minutes. To remove RNA complementary to the cDNA, 1 μl (2 units) of *E. coli* RNase H was added and incubated at 37° C. for 20 minutes. The resulting mixtures were transferred to 0.5 ml tubes and stored at −20° C.

[0284] Oligonucleotide primers were designed based upon the *P. obesus* sequence using Primer Express software (version 1.5), which was obtained at the http address docs.appliedbiosystems.com/pebiodocs/04303014.pdf. The forward primer had the sequence 5' CCA ACA CTT CCG AGC TTG GA 3' and the reverse primer had the sequence 5' TCC TCC TGA GCT GGG ATC AC 3'. These primers were designed to generate a 65 base pair amplicon. Primers for the endogenous control gene, cyclophilin, were designed based on the *P. obesus* sequence. Primer sequence specificity was confirmed by comparing the primer sequences against the GenBank nucleotide sequence for P2X4 using BLAST. Primers were synthesized at a 40 mmole concentration and purified by using a reverse-phase cartridge (GeneWorks, Australia).

[0285] The ability of the primers to operate in a quantitative PCR process was next determined. A standard curve was generated based upon threshold cycles (Ct=threshold cycle) for serially diluted samples. cDNA was serially diluted from a 1:2 dilution to a 1:16 dilution, and the standard curve included an undiluted sample and a "no template control" (contains no cDNA). A standard curve was also generated using primers specific for the endogenous control gene (cyclophilin). These samples are set-up in duplicate using the following: 12.5 μl of SYBR Green Universal PCR master mix (cat# 4304437, http address docs.appliedbiosystems.com/pebiodocs/00777601.pdf); 2.5 μl of forward primer (1 μM, diluted in nuclease-free water); 2.5 μl of reverse primer (1 μM, diluted in nuclease-free water); 2.0 μl of cDNA (neat or diluted); and 5.5 μl of water (nuclease-free) for a total volume of 25 μl.

[0286] The PCR program recommended for the ABI Prism 7700 procedure was utilized and the baseline was calculated based upon cycles 3 to 15 and the amplification plot was based upon cycles 16 to 40. A threshold level was set following examination of a semi-log view of the plot. The Ct values for each duplicate were examined to ensure they did not differ by more than one Ct unit. The Ct values were eliminated or the experiment was repeated if they differed by more than one Ct unit. Samples were run on an agarose gel to identify product formation and whether or not primer-dimers or non-specific priming occurred. While the primer concentration could have been optimized if required, it was determined that 100 nM of each primer (final concentration) was adequate.

[0287] Following the primer efficiency determination, a real time PCR run was executed. The conditions utilized were as described above except that cDNAs were diluted 1:8 and products were not confirmed on an agarose gel. Final values were then calculated using the relation $2^{-\Delta Ct}$, where ΔCt is Ct of cyclophilin subtracted from Ct of the gene of interest. Gene expression values were calculated as arbitrary units, and Ct values for cyclophilin in treated samples (e.g., in fasted tissues) were further examined to determine whether endogenous control of gene expression was altered. This analysis yielded quantified and standardized gene expression values for the amount of cDNA in each reaction.

Example 8

In Vitro Production of P2X4 Polypeptide And P2X4 Polypeptide Variants

[0288] P2X4 cDNA is cloned into a pIVEX 2.3-MCS vector (Roche Biochem) using a directional cloning method. A P2X4 cDNA insert is prepared using PCR with forward and reverse primers having 5' restriction site tags (in frame) and 5-6 additional nucleotides in addition to 3' gene-specific portions, the latter of which is typically about twenty to about twenty-five base pairs in length. A Sal I restriction site is introduced by the forward primer and a Sma I restriction site is introduced by the reverse primer. The ends of P2X4 PCR products are cut with the corresponding restriction enzymes (i.e., Sal I and Sma I) and the products are gel-purified. The pIVEX 2.3-MCS vector is linearized using the same restriction enzymes, and the fragment with the correct sized fragment is isolated by gel-purification. Purified P2X4 PCR product is ligated into the linearized pIVEX 2.3-MCS vector and *E. coli* cells transformed for plasmid amplification. The newly constructed expression vector is verified by restriction mapping and used for protein production.

[0289] *E. coli* lysate is reconstituted with 0.25 ml of Reconstitution Buffer, the Reaction Mix is reconstituted with 0.8 ml of Reconstitution Buffer; the Feeding Mix is reconstituted with 10.5 ml of Reconstitution Buffer; and the Energy Mix is reconstituted with 0.6 ml of Reconstitution Buffer. 0.5 ml of the Energy Mix was added to the Feeding Mix to obtain the Feeding Solution. 0.75 ml of Reaction Mix, 50 μl of Energy Mix, and 10 μg of the P2X4 template DNA is added to the *E. coli* lysate.

[0290] Using the reaction device (Roche Biochem), 1 ml of the Reaction Solution is loaded into the reaction compartment. The reaction device is turned upside-down and 10 ml of the Feeding Solution is loaded into the feeding compartment. All lids are closed and the reaction device is loaded into the RTS500 instrument. The instrument is run at 30° C. for 24 hours with a stir bar speed of 150 rpm. The pIVEX 2.3 MCS vector includes a nucleotide sequence that encodes six consecutive histidine amino acids on the C-terminal end of the P2X4 polypeptide for the purpose of protein purification. P2X4 polypeptide is purified by contacting the contents of reaction device with resin modified with Ni^{2+} ions. P2X4 polypeptide is eluted from the resin with a solution containing free Ni^{2+} ions.

Example 9

Cellular Production of P2X4 Polypeptide and P2X4 Polypeptide Variants

[0291] P2X4 nucleic acids and P2X4 nucleic acid variants are cloned into DNA plasmids having phage recombination

cites and P2X4 polypeptides and polypeptide variants are expressed therefrom in a variety of host cells. X phage genomic DNA contains short sequences known as attP sites, and *E. coli* genomic DNA contains unique, short sequences known as attB sites. These regions share homology, allowing for integration of phage DNA into *E. coli* via directional, site-specific recombination using the phage protein Int and the *E. coli* protein IHF. Integration produces two new att sites, L and R, which flank the inserted prophage DNA. Phage excision from *E. coli* genomic DNA can also be accomplished using these two proteins with the addition of a second phage protein, Xis. DNA vectors have been produced where the integration/excision process is modified to allow for the directional integration or excision of a target DNA fragment into a backbone vector in a rapid in vitro reaction (Gateway™ Technology (Invitrogen, Inc.)).

[0292] A first step is to transfer the P2X4 nucleic acid insert into a shuttle vector that contains attL sites surrounding the negative selection gene, ccdB (e.g. pENTER vector, Invitrogen, Inc.). This transfer process is accomplished by digesting the P2X4 nucleic acid from a DNA vector used for sequencing, and to ligate it into the multicloning site of the shuttle vector, which will place it between the two attL sites while removing the negative selection gene ccdB. A second method is to amplify the P2X4 nucleic acid by the polymerase chain reaction (PCR) with primers containing attB sites. The amplified fragment then is integrated into the shuttle vector using Int and IHF. A third method is to utilize a topoisomerase-mediated process, in which the P2X4 nucleic acid is amplified via PCR using gene-specific primers with the 5' upstream primer containing an additional CACC sequence (e.g., TOPO® expression kit (Invitrogen, Inc.)). In conjunction with Topoisomerase I, the PCR amplified fragment can be cloned into the shuttle vector via the attL sites in the correct orientation.

[0293] Once the P2X4 nucleic acid is transferred into the shuttle vector, it can be cloned into an expression vector having attR sites. Several vectors containing attR sites for expression of P2X4 polypeptide as a native polypeptide, N-fusion polypeptide, and C-fusion polypeptides are commercially available (e.g., pDEST (Invitrogen, Inc.)), and any vector can be converted into an expression vector for receiving a P2X4 nucleic acid from the shuttle vector by introducing an insert having an attR site flanked by an antibiotic resistant gene for selection using the standard methods described above. Transfer of the P2X4 nucleic acid from the shuttle vector is accomplished by directional recombination using Int, 1HF, and Xis (LR clonase). Then the desired sequence can be transferred to an expression vector by carrying out a one hour incubation at room temperature with Int, IHF, and Xis, a ten minute incubation at 37° C. with proteinase K, transforming bacteria and allowing expression for one hour, and then plating on selective media. Generally, 90% cloning efficiency is achieved by this method. Examples of expression vectors are pDEST 14 bacterial expression vector with att7 promoter, pDEST 15 bacterial expression vector with a T7 promoter and a N-terminal GST tag, pDEST 17 bacterial vector with a T7 promoter and a N-terminal polyhistidine affinity tag, and pDEST 12.2 mammalian expression vector with a CMV promoter and neo resistance gene. These expression vectors or others like them are transformed or transfected into cells for expression of the P2X4 polypeptide or polypeptide variants. These expression vectors often are transfected, for example, into murine-transformed a adipocyte cell line 3T3-L1, (ATCC), human embryonic kidney cell line 293, and rat cardiomyocyte cell line H9C2.

Example 10

Cellular Assay for Screening P2X4 Interacting Fat Reduction Drug Candidates

[0294] P2X4 polypeptide or P2X4 polypeptide variants are contacted with a test molecule and an interaction between P2X4 and the test molecule is monitored. Molecules that interact with P2X4 polypeptide or P2X4 polypeptide variants then are characterized further for an effect upon obesity.

[0295] Where the P2X4 polypeptide or P2X4 polypeptide variant is expressed in cell membranes, an interaction between the molecule and the P2X4 polypeptide or P2X4 polypeptide variant is determined by observing an effect of the molecule on calcium uptake and/or levels of cAMP and/or inositol phosphate. Test molecules are selected from libraries of molecules, which can include antagonists of P2X receptors (e.g., suramin, reactive blue 2, PPADS, and others) and agonists of P2X receptors. For example, test molecules may be selected from nucleotides or nucleotide analogs, such as adenosine, ADP, AMP, α,β-methylene-ATP, β,γ-methylene-ATP, UTP, 2-MeSATP, and others.

[0296] Cells that Express P2X4 Polypeptide or P2X4 Polypeptide Variants

[0297] Cells that express P2X4 polypeptides or polypeptide variants (described in Example 6) may be utilized in cellular assays for screening molecules that reduce fat deposition. Also, a myocyte cellular assay described for example in WO 99/23214 can be used to screen for that reduce fat deposition molecules. Regardless of which cell expresses the P2X4 polypeptide or polypeptide variant, the effect of the molecule on P2X4 function is assessed by monitoring calcium uptake and cyclic AMP and inositol phosphate level in the cells, for example. For the myocyte assay, myocytes are cultured by obtaining ventricular cells from chick embryos and culturing them for 14 days in ovo as described in Barry et al., *J. Physiol.*, 325: 243-260 (1982); Liang et al., *Circ. Res.*, 76: 242-251 (1995). After neutralizing trypsin with a medium comprising horse serum, the ventricular cells are centrifuged and resuspended in a culture medium comprising 6% (v/v) fetal bovine serum, 40% (v/v) Medium 199 (GIBCO, Grand Island, N.Y.), 0.1% (w/v each) penicillin/streptomycin, and Hank's salts. The cultured ventricular cells are plated at a density of 400,000 cells per milliliter and cultivated in a humidified 1:19 (vol:vol) CO_2:air atmosphere at 37° C. Cultured cardiac myocytes are grown to confluence on day 3 in culture and exhibit rhythmic spontaneous contraction.

[0298] Calcium Ion Uptake into Cells

[0299] Determination of ^{45}Ca uptake is performed as described in Liang et al., *J. Biol. Chem.*, 271: 18678-18685 (1996). Cultures are incubated with L-(3,4,5-^{3}H, N)-leucine (152.2 Curies per millimole) for 24 hours prior to ^{45}Ca uptake. Incorporation of ^{3}H-leucine into the cellular protein can permit normalization of ^{45}Ca content to milligrams of cell protein by determining 3H content, using known methods. Following exposure to ^{45}Ca, cells then may be washed

free of ^{45}Ca by rinsing the cells four times with ice-cold buffer containing 5 millimolar HEPES buffer, 1 millimolar $CaCl_2$, 4 millimolar KCl, 0.5 millimolar $MgCl_2$, 142 millimolar NaCl, and 1 millimolar lanthanum, at pH 7.35. This washing procedure removes more than 99% of an extracellular marker, ^{51}Cr-EDTA, and substantially removes extracellular ^{45}Ca. Uptake of ^{45}Ca is quantified for about 90 seconds. Cells are solubilized for 2 hours in a solution comprising 1% (w/v) sodium dodecyl sulfate and 10 millimolar sodium borate. Aliquots of the solution containing solubilized cells is assayed for radioactivity and protein content.

[0300] Also, calcium influx can be measured by monitoring a fluorescent signal generated by a compound which binds calcium. A compound for monitoring calcium is normally injected into cells before a modulator is administered, and then fluorescence is measured by standard imaging or photometry procedures at an emission wavelength before and after administration of the modulator. Examples of fluorescent compounds useful for monitoring calcium influx in cells include Quin-2 (ex=339 nm, em=400 nm), Fluo-3 (ex=488 nm, em=530 nm), Fura-Red (ex=488 nm, em=660 nm), Calcium Green (ex=488 nm, em=530 nm), Calcium Orange (ex=550 nm, em=580 nm), Calcium Crimson (ex=593 nm, em=615 nm), Rhod-2 (ex=550 nm, em=575 nm), and Fura-2 (ex_1=340 nm, ex_2=380 nm, em=510 nm), which are available from Intracellular Imaging, Inc. (see e.g., http address www.intracellular.com for structures of compounds and methods of use; abbreviations are "em" for emission wavelength, "ex" for excitation wavelength, "ex," for first excitation wavelength, and "ex_2" for second excitation wavelength).

[0301] Determination of cAMP Level in Cells

[0302] cAMP is extracted by addition of one-tenth of the aliquot volume of 1 normal HCl to the media and boiling the mixture for 10 minutes. Extracted cAMP is assayed according to a radioimmunoassay method described in Liang et al., *Circ. Res.*, 76: 242-251 (1995), Amersham, Arlington Heights, Ill. The effect of P2 receptor agonist on cAMP accumulation is monitored for at least 10 minutes to determine if the effect is linear, after which time the concentration of cAMP is assessed in the sample.

[0303] Determination of Inositol Phosphate Levels in Cells

[0304] Inositol phosphates are monitored according to a method reported by Berridge et al., *Biochem. J.*, 212: 473-482 (1983), as modified by Barnett et al., *Biochem. J.*, 271: 437-442 (1990). For myocytes, cells are pre-incubated with 5 millicuries per milliliter ^{3}H-myo-inositol for 24 hours, washed with Dulbecco's modified Eagle's medium comprising 15 millimolar LiCl (DMEM-Li), and incubated in DMEM-Li for 10 minutes at 37° C. The myocytes then are exposed to one or more test compounds.

[0305] Following this exposure, myocytes are subjected to extraction using 1 milliliter of a solvent comprising a 1:2:0.05 (vol:vol:vol) mixture of chloroform:methanol:HCl to remove inositol phosphates from the myocytes. The solvent comprising inositol phosphates is applied to an anion exchange column (AGx8 resin, formate form, 1 milliliter bed volume) and $InsP_1$, $InsP_2$, and $InsP_3$ are eluted sequentially using a solution comprising 100 millimolar formic acid and 200 millimolar ammonium formate, a solution comprising 100 millimolar formic acid and 600 millimolar ammonium formate, and a solution comprising 100 millimolar formic acid and 1 M ammonium formate, respectively. The anion exchange column is calibrated with each inositol phosphate standard to confirm complete separation of $InsP_1$, $InsP_2$, and $InsP_3$. Recovery of each inositol phosphate often is greater than 95%.

[0306] The effect of ATP receptor agonists on inositol-1, 4,5-trisphosphate ($InsP_3$) level is quantified using an $InsP_3$ radioreceptor assay. Growth media in which ventricular cells are grown can be replaced with a solution comprising HEPES buffer, 1 millimolar $CaCl_2$, 4 millimolar KCl, and 0.5 millimolar $MgCl_2$ at pH 7.35. The cells then are exposed to ATP. The reaction can be terminated by addition of 0.2 volumes of ice-cold trichloroacetic acid (TCA), which is removed by extraction with a solution comprising TCTFE (1,1,2-trichloro-1,2,2-trifluoroethane)-trioctylamine. $InsP_3$ in the aqueous phase may be determined by competition with ^{3}H-$InsP_3$ for binding to the $InsP_3$ receptor supplied as a part of a kit (Dupont, Boston, Mass.; Liang et al., *J. Biol. Chem.*, 271: 18678-18685 (1996)).

[0307] Sources for Materials

[0308] Embryonic chick eggs can be obtained from Spafas Inc. (Storrs, Conn.). A cAMP radioimmunoassay kit can be obtained from Amersham (Arlington Heights, Ill.). ^{3}H-leucine, ^{3}H-myoinositol, an $InSP_3$ radioreceptor assay kit, and ^{45}Ca may be obtained from Dupont (Boston, Mass.). Adenosine, ADP, AMP, α,β-methylene-ATP, β,γ-methylene-ATP, and UTP may be obtained from Sigma Chemical Co. (St. Louis, Mo.). 2-MeSATP may be obtained from Research Biochemical International (Natick, Mass.). U-73122, which is also referred to as 1(6-((17b-3-methoxyestra-1,3,5(10)-trien-17-yl)amino)hexyl)-1H-pyrrole-2,5-dione, and U-73343, also referred to as 1(6-((17b-3-methoxyestra-1,3,5(10)-trien-17-yl)amino)hexyl)-2,5-pyrrolidine-dione may be obtained from BIOMOL (Plymouth Meeting, Pa.).

Example 11

In Vivo Assay for Screening Fat Reduction Drug Candidates

[0309] Test compounds are screened for fat reduction activity by administering molecules which interact with P2X4 to Israeli sand rats (*P. obsesus*), which is an accepted in vivo model for obesity, and observing the effect of the molecule on such parameters as weight, dimensions, and/or fat content. Molecules may be administered to obese animals and/or non-obese animals. These animals are grouped into four sets (Table 9), where group D animals have high morbidity and are not typically used in studies.

TABLE 9

Group	Phenotype	Plasma glucose/Insulin
Group A (27%)	Healthy	Normoglycemic/normoinsulinemic
Group B (35%)	Healthy/Obese	Normoglycemic/hyperinsulinemic
Group C (30%)	Diabetic/Obese	Hyperglycemic/hyperinsulinemic
Group D (8%)	Diabetic/Obese	Hyperglycemic/hypoinsulinemic

[0310] The Israeli sand rat is maintained on an ad libitum diet of a standard lab chow that is high in energy. This

polygenic animal displays in response to this diet a range of body weights, plasma insulin and blood glucose levels. Normally, eight controlled animals and eight treated animals are included for groups A, B and C, giving a total of 48 animals for each study.

[0311] The test molecule is delivered to the animals by intraperitoneal injection; intravenous injection; intragastrical administration, in which case twice as many animals per group should be used since the method of administration is more stressful and leads to a higher motility rate; continuous infusion using an osmotic pump; and orally ad libitum, which is the least stressful as the test molecule is added to food and the amount of consumed is measured. Often DMSO or water is used as a vehicle accompanying the test molecule and 10 μg to 1000 μg of test molecule per kilogram of the animal is typically administered.

[0312] The length of the study is typically one to seven days. During the study, several parameters are measured, including body weight (daily measurements); food intake (daily measurements); blood glucose levels (before and after the study); plasma insulin levels (before and after the study); circulating blood metabolites such as leptin, cortisol, triglycerides and free fatty acids (before and after the study); percent body fat (weighing fat pads at the end of the study); quantification of gene expression in tissues such as the pancreas, mesenteric fat, stomach and small intestine (at the end of the study); and measurements of P2X4 activity in tissues such as pancreas, mesenteric fact, stomach, and small intestine using methods described in Example 7 (before and/or after the study). Animals are sacrificed by anaesthetic overdose and tissues are harvested and rapidly frozen. RNA is extracted from half of each harvested tissue and P2X4 polypeptide extracts are sometimes generated from the other half.

[0313] Modifications may be made to the foregoing without departing from the basic aspects of the invention. Although the invention has been described in substantial detail with reference to one or more specific embodiments, those of skill in the art will recognize that changes may be made to the embodiments specifically disclosed in this application, yet these modifications and improvements are within the scope and spirit of the invention, as set forth in the claims which follow. All publications or patent documents cited in this specification are incorporated herein by reference as if each such publication or document was specifically and individually indicated to be incorporated herein by reference.

[0314] Citation of the above publications or documents is not intended as an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents. U.S. patents and other publications referenced herein are hereby incorporated by reference.

What is claimed is:

1. A method for diagnosing a predisposition to fat deposition in a subject, which comprises detecting the presence or absence of a polymorphic variation associated with fat deposition at a polymorphic site in a P2X4 nucleotide sequence in a nucleic acid sample from a subject, wherein the P2X4 nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of:

(a) the nucleotide sequence of SEQ ID NO:1;

(b) a nucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; and

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising the polymorphic site;

whereby the presence of the polymorphic variation is indicative of a predisposition to fat deposition in the subject.

2. The method of claim 1, which further comprises obtaining the nucleic acid sample from the subject.

3. The method of claim 1, wherein the polymorphic variation is an adenine at position 11030 of SEQ ID NO:1.

4. The method of claim 3, wherein the polymorphic variation is in linkage disequilibrium with the adenine at position 11030 of SEQ ID NO:1.

5. The method of claim 1, wherein the polymorphic variation is a thymine at position 15847 of SEQ ID NO:1.

6. The method of claim 5, wherein the polymorphic variation is in linkage disequilibrium with the thymine at position 15847 of SEQ ID NO:1.

7. The method of claim 1, wherein the polymorphic variation is a cytosine at position 17338 of SEQ ID NO:1.

8. The method of claim 7, wherein the polymorphic variation is in linkage disequilibrium with the cytosine at position 17338 of SEQ ID NO:1.

9. The method of claim 1, wherein the polymorphic variation is a GGGCCCC nucleic acid insert between positions 2878 and 2891 in SEQ ID NO:1.

10. The method of claim 9, wherein the polymorphic variation is in linkage disequilibrium with the GGGCCCC nucleic acid insert between positions 2878 and 2891 of SEQ ID NO:1.

11. The method of claim 1, wherein detecting the presence or absence of a polymorphic variation comprises:

hybridizing an oligonucleotide to the nucleic acid sample, wherein the oligonucleotide is complementary to the P2X4 nucleotide sequence and hybridizes to a region of the P2X4 nucleotide sequence that is adjacent to the polymorphic variation;

extending the oligonucleotide in the presence of one or more nucleotides, yielding extension products; and

detecting the presence or absence of the polymorphic variation in the extension products.

12. The method of claim 11, wherein the oligonucleotide is selected from the group consisting of AAAAAAGACCGAACAC (SEQ ID NO: 5), CATTAGCTCGGTCTCC (SEQ ID NO: 6), GCTTTCAACGGGTC (SEQ ID NO: 7), TAGCTGGGATTACAGAC (SEQ ID NO: 8), CATGTCCTGGAAAC (SEQ ID NO: 9), GCCTTCTGCCTTGTA (SEQ ID NO: 10), CCTCAAAAAAAGACCGAACAC (SEQ ID NO: 11), GCGTAGCTTTCAACGGGTC (SEQ ID NO: 12), ACGGCCATGTCCTGGAAAC (SEQ ID NO: 13), GAAGAAAAGACACAGAGACAC (SEQ ID NO: 14), TGGCAGGATGTTTCTCCTGT (SEQ ID NO: 15), AGGATCCGGAAAGAGTTTCC (SEQ ID NO: 16), CAGCTTTCTCTACTATATTCATGTAA (SEQ ID NO: 17), GCTCATTAGCTCGGTCTCC (SEQ ID NO: 18),

GGAAAAACACGAAAGGCTTAGTT (SEQ ID NO: 19), TCAGGCATGTGATCAGTTAGAGCC (SEQ ID NO: 20), and CCTTCCCACTGAGCACATC (SEQ ID NO: 21).

13. The method of claim 1, wherein the fat deposition is central fat deposition in the subject.

14. The method of claim 1, wherein the subject is a human.

15. A method for diagnosing a predisposition to leanness in a subject, which comprises detecting the presence or absence of a polymorphic variation associated with leanness at a polymorphic site in a P2X4 nucleotide sequence in a nucleic acid sample from a subject, wherein the P2X4 nucleotide sequence is selected from the group consisting of:

(a) the nucleotide sequence of SEQ ID NO:1;

(b) a nucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; and

(d) a fragment of a nucleotide sequence of (i), (ii), or (iii) comprising the polymorphic site;

whereby the presence of the polymorphic variation is indicative of leanness in the subject.

16. The method of claim 15, wherein the polymorphic variation is selected from the group consisting of a thymine at position 11030, a thymine at position 15847, a thymine at position 17338, and the sequence GGGCCCCGGGCCCC between positions 2878 and 2891 in SEQ ID NO:1.

17. The method of claim 15, wherein the polymorphic variation is a thymine at position 11030, a thymine at position 15847, a thymine at position 17338, an adenine at position 21708, and a thymine at position 22713 of SEQ ID NO:1.

18. A method for identifying a polymorphic variation associated with fat deposition proximal to an incident polymorphic variation associated with fat deposition, which comprises:

identifying a polymorphic variant proximal to the incident polymorphic variant associated with fat deposition, wherein the incident polymorphic variant is in a P2X4 nucleotide sequence and the P2X4 nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of:

(a) a polynucleotide sequence set forth in SEQ ID NO: 1;

(b) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence encoded by a nucleotide sequence set forth as SEQ ID NO: 1; or

(c) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence that is 90% identical to an amino acid sequence encoded by a nucleotide sequence set forth in SEQ ID NO: 1; and

determining the presence or absence of an association of the proximal polymorphic variant with fat deposition.

19. The method of claim 18, wherein the first polymorphic variant is located at position 2878, 11030, 15847, or 17338 of SEQ ID NO: 1.

20. The method of claim 18, wherein the proximal polymorphic variant is within a region spanning about 5 kb 5' of the incident polymorphic variant and about 5 kb 3' of the incident polymorphic variant.

21. The method of claim 18, which further comprises determining if the proximal polymorphic variant is in linkage disequilibrium with the incident polymorphic variant.

22. The method of claim 18, which further comprises identifying a second polymorphic variant proximal to a proximal polymorphic variant of claim 18 associated with fat deposition and determining if the second polymorphic variant is associated with fat deposition.

23. The method of claim 22, wherein the second proximal polymorphic variant is within a region spanning about 5 kb 5' of the incident polymorphic variant and about 5 kb 3' of the proximal polymorphic variant associated with fat deposition.

24. A method for diagnosing a predisposition to non-insulin dependent diabetes mellitus (NIDDM) in a subject, which comprises detecting the presence or absence of a polymorphic variation associated with NIDDM at a polymorphic site in a P2X4 nucleotide sequence in a nucleic acid sample from a subject, wherein the P2X4 nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of:

(a) the nucleotide sequence of SEQ ID NO:1;

(b) a nucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; and

(d) a fragment of a nucleotide sequence of (i), (ii), or (iii) comprising the polymorphic site;

whereby the presence of the polymorphic variation is indicative of a predisposition to NIDDM in the subject.

25. The method of claim 18, wherein the polymorphic variation is a thymine at position 15847 of SEQ ID NO:1.

26. The method of claim 19, wherein the polymorphic variation is in linkage disequilibrium with the thymine at position 15847 of SEQ ID NO: 1.

27. A method for identifying a polymorphic variation associated with NIDDM proximal to an incident polymorphic variation associated with NIDDM, which comprises:

identifying a polymorphic variant proximal to the incident polymorphic variant associated with NIDDM, wherein the incident polymorphic variant is in a P2X4 nucleotide sequence and the P2X4 nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of:

(a) a polynucleotide sequence set forth in SEQ ID NO: 1;

(b) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence encoded by a nucleotide sequence set forth as SEQ ID NO: 1; or

(c) a polynucleotide sequence that encodes a polypeptide having an amino acid sequence that is 90% identical to an amino acid sequence encoded by a nucleotide sequence set forth in SEQ ID NO: 1; and

determining the presence or absence of an association of the proximal polymorphic variant with NIDDM.

28. The method of claim 27, wherein the first polymorphic variant is a thymine at position 15847 of SEQ ID NO: 1.

29. The method of claim 27, wherein the proximal polymorphic variant is within a region spanning about 5 kb 5' of the incident polymorphic variant and about 5 kb 3' of the incident polymorphic variant.

30. The method of claim 27, which further comprises determining if the proximal polymorphic variant is in linkage disequilibrium with the incident polymorphic variant.

31. The method of claim 27, which further comprises identifying a second polymorphic variant proximal to a proximal polymorphic variant of claim 27 associated with NIDDM and determining if the second polymorphic variant is associated with NIDDM.

32. The method of claim 31, wherein the second proximal polymorphic variant is within a region spanning about 5 kb 5' of the incident polymorphic variant and about 5 kb 3' of the proximal polymorphic variant associated with NIDDM.

33. An isolated nucleic acid which comprises a P2X4 nucleotide sequence selected from the group consisting of:

(a) the nucleotide sequence of SEQ ID NO:1;

(b) a nucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3;

(d) a fragment of a nucleotide sequence of (a), (b), or (c) comprising a polymorphic variation; and

(e) a nucleotide sequence complementary to the nucleotide sequences of (a), (b), (c), or (d);

wherein the P2X4 nucleotide sequence comprises one or more polymorphic variations selected from the group consisting of thymine at position 14744, an adenine at position 14977, a cytosine at position 15046, a thymine at position 21538, a guanine at position 21630, and a single nucleotide deletion at position 25264 of SEQ ID NO:1.

34. An oligonucleotide which comprises a nucleotide sequence complementary to a portion of a P2X4 nucleotide sequence selected from the group consisting of:

(a) the nucleotide sequence of SEQ ID NO:1;

(b) a nucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; and

(c) a nucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3;

wherein the 3' end of the oligonucleotide is adjacent to a polymorphic variation in the P2X4 nucleotide sequence.

35. The oligonucleotide of claim 34, which consists of a nucleotide sequence selected from the group consisting of AAAAAAAGACCGAACAC (SEQ ID NO: 5), CATTAGCTCGGTCTCC (SEQ ID NO: 6), GCTTTCAACGGGTC (SEQ ID NO: 7), TAGCTGGGATTACAGAC (SEQ ID NO: 8), CATGTCCTGGAAAC (SEQ ID NO: 9), GCCTTCTGCCTTGTA (SEQ ID NO: 10), CCTCAAAAAAAGACCGAACAC (SEQ ID NO: 11), GCGTAGCTTTCAACGGGTC (SEQ ID NO: 12), ACGGCCATGTCCTGGAAAC (SEQ ID NO: 13), GAAGAAAAGACACAGAGACAC (SEQ ID NO: 14), TGGCAGGATGTTTCTCCTGT (SEQ ID NO: 15), AGGATCCGGAAAGAGTTTCC (SEQ ID NO: 16), CAGCTTTCTACTATATTCATGTAA (SEQ ID NO: 17), GCTCATTAGCTCGGTCTCC (SEQ ID NO: 18), GGAAAAAACACGAAAGGCTTAGTT (SEQ ID NO: 19), TCAGGCATGTGATCAGTTAGAGCC (SEQ ID NO: 20), and CCTTCCCACTGAGCACATC (SEQ ID NO: 21).

36. A microarray comprising an isolated nucleic acid of claim 33 linked to a solid support.

37. An isolated polypeptide encoded by the isolated nucleic acid sequence of claim 33.

38. A method for identifying a candidate therapeutic for fat reduction, which comprises:

(a) introducing a test molecule to a system which comprises a nucleic acid comprising a P2X4 nucleotide sequence selected from the group consisting of:

(i) the nucleotide sequence of SEQ ID NO:1;

(ii) a nucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3;

(iii) a nucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; and

(iv) a fragment of a nucleotide sequence of (i), (ii), or (iii); or

introducing a test molecule to a system which comprises a protein encoded by a nucleotide sequence of (i), (ii), (iii), or (iv); and

(b) determining the presence or absence of an interaction between the test molecule and the nucleic acid or protein,

whereby the presence of an interaction between the test molecule and the nucleic acid or protein identifies the test molecule as a candidate therapeutic for fat reduction.

39. The method of claim 38, wherein the system is an animal.

40. The method of claim 38, wherein the system is a cell.

41. The method of claim 38, wherein the P2X4 nucleotide sequence comprises one or more polymorphic variations selected from the group consisting of a an adenine at position 11030, a thymine at position 11030, a thymine at position 15847, a cytosine at position 15847, a cytosine at position 17338, and a thymine at position 17338.

42. A method for reducing fat deposition in a subject, which comprises administering a candidate therapeutic of claim 38 to a subject in need thereof, whereby the candidate therapeutic reduces fat deposition in the subject.

43. A method for reducing fat deposition in a subject, which comprises contacting a P2X4 nucleic acid with one or more cells of a subject in need thereof, wherein the P2X4 nucleic acid comprises a nucleotide sequence selected from the group consisting of:

(a) the nucleotide sequence of SEQ ID NO:1;

(b) a nucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; and

(d) a fragment of a nucleotide sequence of (i), (ii), or (iii);

whereby contacting the one or more cells of the subject with the P2X4 nucleic acid reduces fat deposition.

44. A method for reducing fat deposition in a subject, which comprises contacting a P2X4 protein with one or more cells of a subject in need thereof, wherein the P2X4 protein is encoded by a P2X4 nucleotide sequence which comprises a polynucleotide sequence selected from the group consisting of:

(a) the polynucleotide sequence of SEQ ID NO:1;

(b) a polynucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3;

(c) a polynucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; and

(d) a fragment of a polynucleotide sequence of (i), (ii), or (iii);

whereby contacting the one or more cells of the subject with the P2X4 protein reduces fat deposition.

45. A method for reducing fat deposition in a subject, which comprises:

detecting the presence or absence of a polymorphic variant associated with fat deposition in a P2X4 nucleotide sequence in a nucleic acid sample from a subject, wherein the P2X4 nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of:

(a) the polynucleotide sequence of SEQ ID NO:1;

(b) a polynucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3;

(c) a polynucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; and

(d) a fragment of a polynucleotide sequence of (i), (ii), or (iii); and

administering a treatment that reduces fat deposition to a subject from whom the sample originated where the presence of a polymorphic variation associated with fat reduction is detected in the P2X4 nucleotide sequence.

46. The method of claim 45, wherein the polymorphic variation is selected from the group consisting of an adenine at position 11030, a thymine at position 15847, and a cytosine at position 17338.

47. The method of claim 45, wherein the treatment is one or more selected from the group consisting of an appetite suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise regimen, a dietary regimen, psychological consoling, psychotherapy, and a psychotherapeutic.

48. A method for identifying a candidate therapeutic for alleviating NIDDM, which comprises:

(a) introducing a test molecule to a system which comprises a nucleic acid comprising a P2X4 nucleotide sequence selected from the group consisting of:

(i) the nucleotide sequence of SEQ ID NO: 1;

(ii) a nucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3;

(iii) a nucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; and

(iv) a fragment of a nucleotide sequence of (i), (ii), or (iii); or

introducing a test molecule to a system which comprises a protein encoded by a nucleotide sequence of (i), (ii), (iii), or (iv); and

(b) determining the presence or absence of an interaction between the test molecule and the nucleic acid or protein,

whereby the presence of an interaction between the test molecule and the nucleic acid or protein identifies the test molecule as a candidate therapeutic for treating NIDDM.

49. The method of claim 48, wherein the system is an animal.

50. The method of claim 48, wherein the system is a cell.

51. The method of claim 48, wherein the P2X4 nucleotide sequence comprises a thymine at position 15847 of SEQ ID NO:1.

52. A method for treating NIDDM in a subject, which comprises administering a candidate therapeutic of claim 48 to the subject in need thereof, whereby the candidate therapeutic treats NIDDM in the subject.

53. A method for alleviating NIDDM in a subject, which comprises contacting a P2X4 nucleic acid with one or more cells of a subject in need thereof, wherein the P2X4 nucleic acid comprises a nucleotide sequence selected from the group consisting of:

(a) the nucleotide sequence of SEQ ID NO:1;

(b) a nucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3;

(c) a nucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; and

(d) a fragment of a nucleotide sequence of (i), (ii), or (iii);

whereby contacting the one or more cells of the subject with the P2X4 nucleic acid alleviates NIDDM.

54. A method for alleviating NIDDM in a subject, which comprises contacting a P2X4 protein with one or more cells of a subject in need thereof, wherein the P2X4 protein is encoded by a P2X4 nucleotide sequence which comprises a polynucleotide sequence selected from the group consisting of:

(a) the polynucleotide sequence of SEQ ID NO:1;

(b) a polynucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3;

(c) a polynucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; and

(d) a fragment of a polynucleotide sequence of (i), (ii), or (iii);

whereby contacting the one or more cells of the subject with the P2X4 protein alleviates NIDDM.

55. A method for alleviating NIDDM in a subject, which comprises:

detecting the presence or absence of a polymorphic variant associated with NIDDM in a P2X4 nucleotide sequence in a nucleic acid sample from a subject, wherein the P2X4 nucleotide sequence comprises a polynucleotide sequence selected from the group consisting of:

(a) the polynucleotide sequence of SEQ ID NO:1;

(b) a polynucleotide sequence which encodes a polypeptide consisting of the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3;

(c) a polynucleotide sequence which encodes a polypeptide that is 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3; and

(d) a fragment of a polynucleotide sequence of (i), (ii), or (iii); and

administering a treatment that alleviates NIDDM to a subject from whom the sample originated where the presence of a polymorphic variation associated with NIDDM is detected in the P2X4 nucleotide sequence.

56. The method of claim 55, wherein the polymorphic variation is a thymine at position 15847 of SEQ ID NO:1.

57. The method of claim 55, wherein the treatment is one or more selected from the group consisting of insulin, a hypoglycemic, a starch blocker, a liver glucose regulating agent, an insulin sensitizer, a glucose level monitoring regimen, dietary counseling, and a dietary regimen for managing blood glucose levels.

* * * * *